FORM PTO-1390 U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE		ATTORNEY'S DOCKET NUMBER
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371		**PT-1086 USN**
		U.S. APPLICATION NO. (If known, see 37 CFR 1.5) TO BE ASSIGNED 10/089644
INTERNATIONAL APPLICATION NO. **PCT/US00/26085**	INTERNATIONAL FILING DATE **22 September 2000**	PRIORITY DATE CLAIMED **28 September 1999**

TITLE OF INVENTION
SECRETORY MOLECULES

APPLICANT(S) FOR DO/EO/US
HODGSON, David; LINCOLN, Stephen, E.; RUSSO, Frank D.; SPIRO, Peter A.; BANVILLE, Steven C.; BRATCHER, Shawn, R.; DUFOUR, Gerard, E.; COHEN, Howard J.; ROSEN, Bruce H.; SHAH, Purvi; CHALUP, Michael, S.; HILLMAN, Jennifer L.; JONES, Anissa L.; YU, Jimmy Y.; GREENAWALT, Lila B.; PANZER, Scott R.; ROSEBERRY, Ann M.; WRIGHT, Rachel J.; CHEN, Wensheng; LIU, Tommy F.; YAP, Pierre E.; STOCKDREHER, Theresa K.; AMSHEY, Stefan; FONG, Willy T.

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is the **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☐ This is an express request to promptly begin national examination procedures (35 U.S.C. 371 (f)).
4. ☐ The US has been elected by the expiration of 19 months from the priority date (PCT Article 31).
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. ☐ is attached hereto (required only if not communicated by the International Bureau)
   b. ☐ has been communicated by the International Bureau.
   c. ☒ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
7. ☐ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. ☐ are attached hereto (required only if not communicated by the International Bureau).
   b. ☐ have been communicated by the International Bureau.
   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   d. ☐ have not been made and will not be made.
   e. ☐ attached hereto Article 34 Amendment
8. ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☒ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ An English language translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 16 below concern document(s) or information included:

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☒ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.27 and 3.31 is included.
13. ☐ A FIRST preliminary amendment.
    ☐ A SECOND or SUBSEQUENT preliminary amendment.
14. ☐ A substitute specification.
15. ☐ A change of power of attorney and/or address letter.
16. ☒ Other items or information:

1) Transmittal Letter (2 pp, in duplicate)
2) Return Postcard
3) Express Mail Label No.: **EL 856 149 265 US**
4) Sequence Listing on Diskette
5) Sequence Listing Statement
6) Copy of International Search Report (PCT/ISA/210)

U.S. APPLICATION NO. (if known, see 37 CFR 1.5) TO BE ASSIGNED	INTERNATIONAL APPLICATION NO.: PCT/US00/26085	ATTORNEY'S DOCKET NUMBER PT-1086 USN

17. ☒ The following fees are submitted: **BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(5):** Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO.....$1000.00 ☐International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO..$860.00 International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO.....$710.00 ☒**International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4).....$710.00** ☐International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4).....$100.00					
ENTER APPROPRIATE BASIC FEE AMOUNT =				$710.00	
Surcharge of **$130.00** for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).				$	
CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total Claims	19 =	0	X **$ 18.00**	$	
Independent Claims	1 =	0	X **$ 80.00**	$	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			**+ $270.00**	$	
TOTAL OF ABOVE CALCULATIONS =				$	
☐ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.				$	
= **SUBTOTAL**				$710.00	
Processing fee of **$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest clailmed priority date (37 CFR 1492(f)). +				$	
TOTAL NATIONAL FEE =				$710.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by the appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +					
TOTAL FEES ENCLOSED =				$710.00	
				Amount to be Refunded:	$
				Charged:	$

a. ☐ A check in the amount of $_____ to cover the above fees is enclosed.
b. ☒ Please charge my Deposit Account No. 09-0108 in the amount of $710.00 to cover the above fees.
c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 09-0108. A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

INCYTE GENOMICS, INC.
3160 Porter Drive
Palo Alto, CA 94304

SIGNATURE

NAME: Diana Hamlet-Cox

REGISTRATION NUMBER: 33,302

DATE: 27 March 2002

MOLECULES FOR DISEASE DETECTION AND TREATMENT

This application claims the benefit of U.S. Ser. No. 60/156,565 filed September 28, 1999 and U.S. Ser. No. 60/168,197 filed November 30, 1999.

TECHNICAL FIELD

The present invention relates to molecules for disease detection and treatment and to the use of these sequences in the diagnosis, study, prevention, and treatment of diseases associated with, as well as effects of exogenous compounds on, the expression of molecules for disease detection and treatment.

BACKGROUND OF THE INVENTION

The human genome is comprised of thousands of genes, many encoding gene products that function in the maintenance and growth of the various cells and tissues in the body. Aberrant expression or mutations in these genes and their products is the cause of, or is associated with, a variety of human diseases such as cancer and other cell proliferative disorders. The identification of these genes and their products is the basis of an ever-expanding effort to find markers for early detection of diseases, and targets for their prevention and treatment.

For example, cancer represents a type of cell proliferative disorder that affects nearly every tissue in the body. A wide variety of molecules, either aberrantly expressed or mutated, can be the cause of, or involved with, various cancers because tissue growth involves complex and ordered patterns of cell proliferation, cell differentiation, and apoptosis. Cell proliferation must be regulated to maintain both the number of cells and their spatial organization. This regulation depends upon the appropriate expression of proteins which control cell cycle progression in response to extracellular signals such as growth factors and other mitogens, and intracellular cues such as DNA damage or nutrient starvation. Molecules which directly or indirectly modulate cell cycle progression fall into several categories, including growth factors and their receptors, second messenger and signal transduction proteins, oncogene products, tumor-suppressor proteins, and mitosis-promoting factors. Aberrant expression or mutations in any of these gene products can result in cell proliferative disorders such as cancer. Oncogenes are genes generally derived from normal genes that, through abnormal expression or mutation, can effect the transformation of a normal cell to a malignant one (oncogenesis). Oncoproteins, encoded by oncogenes, can affect cell proliferation in a variety of ways and include growth factors, growth factor receptors, intracellular signal transducers, nuclear transcription factors, and cell-cycle control proteins. In contrast, tumor-suppressor genes are involved in inhibiting cell proliferation. Mutations which cause reduced or loss of function in tumor-suppressor genes result in aberrant cell proliferation and cancer. Thus a wide variety of genes

and their products have been found that are associated with cell proliferative disorders such as cancer, but many more may exist that are yet to be discovered.

DNA-based arrays can provide a simple way to explore the expression of a single polymorphic gene or a large number of genes. When the expression of a single gene is explored, DNA-based arrays are employed to detect the expression of specific gene variants. For example, a p53 tumor suppressor gene array is used to determine whether individuals are carrying mutations that predispose them to cancer. A cytochrome p450 gene array is useful to determine whether individuals have one of a number of specific mutations that could result in increased drug metabolism, drug resistance or drug toxicity.

DNA-based array technology is especially relevant for the rapid screening of expression of a large number of genes. There is a growing awareness that gene expression is affected in a global fashion. A genetic predisposition, disease or therapeutic treatment may affect, directly or indirectly, the expression of a large number of genes. In some cases the interactions may be expected, such as when the genes are part of the same signaling pathway. In other cases, such as when the genes participate in separate signaling pathways, the interactions may be totally unexpected. Therefore, DNA-based arrays can be used to investigate how genetic predisposition, disease, or therapeutic treatment affects the expression of a large number of genes.

The discovery of new molecules for disease detection and treatment satisfies a need in the art by providing new compositions which are useful in the diagnosis, study, prevention, and treatment of diseases associated with, as well as effects of exogenous compounds on, the expression of molecules for disease detection and treatment.

SUMMARY OF THE INVENTION

The present invention relates to human disease detection and treatment molecule polynucleotides (mddt) as presented in the Sequence Listing. The mddt uniquely identify genes encoding structural, functional, and regulatory disease detection and treatment molecules.

The invention provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; c) a polynucleotide sequence complementary to a); d) a polynucleotide sequence complementary to b); and e) an RNA equivalent of a) through d). In one alternative, the polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25. In another alternative, the polynucleotide comprises at least 60 contiguous nucleotides of a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; b) a naturally occurring polynucleotide sequence having at least 90%

sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; c) a polynucleotide sequence complementary to a); d) a polynucleotide sequence complementary to b); and e) an RNA equivalent of a) through d). The invention further provides a composition for the detection of expression of disease detection and treatment molecule polynucleotides comprising at least one isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; c) a polynucleotide sequence complementary to a); d) a polynucleotide sequence complementary to b); and e) an RNA equivalent of a) through d); and a detectable label.

The invention also provides a method for detecting a target polynucleotide in a sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; c) a polynucleotide sequence complementary to a); d) a polynucleotide sequence complementary to b); and e) an RNA equivalent of a) through d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide, and b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof. In one alternative, the probe comprises at least 30 contiguous nucleotides. In another alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a recombinant polynucleotide comprising a promoter sequence operably linked to an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; c) a polynucleotide sequence complementary to a); d) a polynucleotide sequence complementary to b); and e) an RNA equivalent of a) through d). In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide. In a further alternative, the invention provides a method for producing a disease detection and treatment molecule polypeptide, the method comprising a) culturing a cell under conditions suitable for expression of the disease detection and treatment molecule polypeptide, wherein said cell is transformed with the recombinant polynucleotide, and b)

recovering the disease detection and treatment molecule polypeptide so expressed.

The invention also provides a purified disease detection and treatment molecule polypeptide (MDDT) encoded by at least one polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25. Additionally, the invention provides an isolated antibody which specifically binds to the disease detection and treatment molecule polypeptide. The invention further provides a method of identifying a test compound which specifically binds to the disease detection and treatment molecule polypeptide, the method comprising the steps of a) providing a test compound; b) combining the disease detection and treatment molecule polypeptide with the test compound for a sufficient time and under suitable conditions for binding; and c) detecting binding of the disease detection and treatment molecule polypeptide to the test compound, thereby identifying the test compound which specifically binds the disease detection and treatment molecule polypeptide.

The invention further provides a microarray wherein at least one element of the microarray is an isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; c) a polynucleotide sequence complementary to a); d) a polynucleotide sequence complementary to b); and e) an RNA equivalent of a) through d). The invention also provides a method for generating a transcript image of a sample which contains polynucleotides. The method comprises a) labeling the polynucleotides of the sample, b) contacting the elements of the microarray with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and c) quantifying the expression of the polynucleotides in the sample.

Additionally, the invention provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; c) a polynucleotide sequence complementary to a); d) a polynucleotide sequence complementary to b); and e) an RNA equivalent of a) through d). The method comprises a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the

group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25; iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv), and alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i-v above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

DESCRIPTION OF THE TABLES

Table 1 shows the sequence identification numbers (SEQ ID NO:s) and template identification numbers (template IDs) corresponding to the polynucleotides of the present invention, along with their GenBank hits (GI Numbers), probability scores, and functional annotations corresponding to the GenBank hits.

Table 2 shows the sequence identification numbers (SEQ ID NO:s) and template identification numbers (template IDs) corresponding to the polynucleotides of the present invention, along with polynucleotide segments of each template sequence as defined by the indicated "start" and "stop" nucleotide positions. The reading frames of the polynucleotide segments and the Pfam hits, Pfam descriptions, and E-values corresponding to the polypeptide domains encoded by the polynucleotide segments are indicated.

Table 3 shows the sequence identification numbers (SEQ ID NO:s) and template identification numbers (template IDs) corresponding to the polynucleotides of the present invention, along with polynucleotide segments of each template sequence as defined by the indicated "start" and "stop" nucleotide positions. The reading frames of the polynucleotide segments are shown, and the polypeptides encoded by the polynucleotide segments constitute either signal peptide (SP) or transmembrane (TM) domains, as indicated.

Table 4A and Table 4B show the sequence identification numbers (SEQ ID NO:s) and

template identification numbers (template IDs) corresponding to the polynucleotides of the present invention, along with component sequence identification numbers (component IDs) corresponding to each template. The component sequences, which were used to assemble the template sequences, are defined by the indicated "start" and "stop" nucleotide positions along each template.

Table 5 shows the tissue distribution profiles for the templates of the invention.

Table 6 summarizes the bioinformatics tools which are useful for analysis of the polynucleotides of the present invention. The first column of Table 6 lists analytical tools, programs, and algorithms, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences).

DETAILED DESCRIPTION OF THE INVENTION

Before the nucleic acid sequences and methods are presented, it is to be understood that this invention is not limited to the particular machines, methods, and materials described. Although particular embodiments are described, machines, methods, and materials similar or equivalent to these embodiments may be used to practice the invention. The preferred machines, methods, and materials set forth are not intended to limit the scope of the invention which is limited only by the appended claims.

The singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. All technical and scientific terms have the meanings commonly understood by one of ordinary skill in the art. All publications are incorporated by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are presented and which might be used in connection with the invention. Nothing in the specification is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

Definitions

As used herein, the lower case "mddt" refers to a nucleic acid sequence, while the upper case "MDDT" refers to an amino acid sequence encoded by mddt. A "full-length" mddt refers to a nucleic acid sequence containing the entire coding region of a gene endogenously expressed in human tissue.

"Adjuvants" are materials such as Freund's adjuvant, mineral gels (aluminum hydroxide), and surface active substances (lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol) which may be administered to increase a host's immunological response.

"Allele" refers to an alternative form of a nucleic acid sequence. Alleles result from a

"mutation," a change or an alternative reading of the genetic code. Any given gene may have none, one, or many allelic forms. Mutations which give rise to alleles include deletions, additions, or substitutions of nucleotides. Each of these changes may occur alone, or in combination with the others, one or more times in a given nucleic acid sequence. The present invention encompasses allelic mddt.

"Amino acid sequence" refers to a peptide, a polypeptide, or a protein of either natural or synthetic origin. The amino acid sequence is not limited to the complete, endogenous amino acid sequence and may be a fragment, epitope, variant, or derivative of a protein expressed by a nucleic acid sequence.

"Amplification" refers to the production of additional copies of a sequence and is carried out using polymerase chain reaction (PCR) technologies well known in the art.

"Antibody" refers to intact molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind MDDT polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or peptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

"Antisense sequence" refers to a sequence capable of specifically hybridizing to a target sequence. The antisense sequence may include DNA, RNA, or any nucleic acid mimic or analog such as peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine.

"Antisense sequence" refers to a sequence capable of specifically hybridizing to a target sequence. The antisense sequence can be DNA, RNA, or any nucleic acid mimic or analog.

"Antisense technology" refers to any technology which relies on the specific hybridization of an antisense sequence to a target sequence.

A "bin" is a portion of computer memory space used by a computer program for storage of data, and bounded in such a manner that data stored in a bin may be retrieved by the program.

"Biologically active" refers to an amino acid sequence having a structural, regulatory, or biochemical function of a naturally occurring amino acid sequence.

"Clone joining" is a process for combining gene bins based upon the bins' containing sequence information from the same clone. The sequences may assemble into a primary gene transcript as well as one or more splice variants.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing (5'-A-G-T-3' pairs with its complement 3'-T-C-A-5').

A "component sequence" is a nucleic acid sequence selected by a computer program such as PHRED and used to assemble a consensus or template sequence from one or more component sequences.

A "consensus sequence" or "template sequence" is a nucleic acid sequence which has been assembled from overlapping sequences, using a computer program for fragment assembly such as the GELVIEW fragment assembly system (Genetics Computer Group (GCG), Madison WI) or using a relational database management system (RDMS).

"Conservative amino acid substitutions" are those substitutions that, when made, least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain.

"Deletion" refers to a change in either a nucleic or amino acid sequence in which at least one nucleotide or amino acid residue, respectively, is absent.

"Derivative" refers to the chemical modification of a nucleic acid sequence, such as by replacement of hydrogen by an alkyl, acyl, amino, hydroxyl, or other group.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

"E-value" refers to the statistical probability that a match between two sequences occurred by chance.

A "fragment" is a unique portion of mddt or MDDT which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 10 to 1000 contiguous amino acid residues or nucleotides. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous amino acid residues or nucleotides in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing and the figures, may be encompassed by the present embodiments.

A fragment of mddt comprises a region of unique polynucleotide sequence that specifically identifies mddt, for example, as distinct from any other sequence in the same genome. A fragment of mddt is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish mddt from related polynucleotide sequences. The precise length of a fragment of mddt and the region of mddt to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of MDDT is encoded by a fragment of mddt. A fragment of MDDT comprises a region of unique amino acid sequence that specifically identifies MDDT. For example, a fragment of MDDT is useful as an immunogenic peptide for the development of antibodies that specifically recognize MDDT. The precise length of a fragment of MDDT and the region of MDDT to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" nucleotide sequence is one containing at least a start site for translation to a protein sequence, followed by an open reading frame and a stop site, and encoding a "full length" polypeptide.

"Hit" refers to a sequence whose annotation will be used to describe a given template. Criteria for selecting the top hit are as follows: if the template has one or more exact nucleic acid matches, the top hit is the exact match with highest percent identity. If the template has no exact matches but has significant protein hits, the top hit is the protein hit with the lowest E-value. If the

template has no significant protein hits, but does have significant non-exact nucleotide hits, the top hit is the nucleotide hit with the lowest E-value.

"Homology" refers to sequence similarity either between a reference nucleic acid sequence and at least a fragment of an mddt or between a reference amino acid sequence and a fragment of an MDDT.

"Hybridization" refers to the process by which a strand of nucleotides anneals with a complementary strand through base pairing. Specific hybridization is an indication that two nucleic acid sequences share a high degree of identity. Specific hybridization complexes form under defined annealing conditions, and remain hybridized after the "washing" step. The defined hybridization conditions include the annealing conditions and the washing step(s), the latter of which is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid probes that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency.

Generally, stringency of hybridization is expressed with reference to the temperature under which the wash step is carried out. Generally, such wash temperatures are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization is well known and can be found in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, or 55°C may be used. SSC concentration may be varied from about 0.2 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, denatured salmon sperm DNA at about 100-200 µg/ml. Useful variations on these conditions will be readily apparent to those skilled in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their resultant proteins.

Other parameters, such as temperature, salt concentration, and detergent concentration may be varied to achieve the desired stringency. Denaturants, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as RNA:DNA hybridizations. Appropriate hybridization conditions are routinely determinable by one of ordinary

skill in the art.

"Immunogenic" describes the potential for a natural, recombinant, or synthetic peptide, epitope, polypeptide, or protein to induce antibody production in appropriate animals, cells, or cell lines.

"Insertion" or "addition" refers to a change in either a nucleic or amino acid sequence in which at least one nucleotide or residue, respectively, is added to the sequence.

"Labeling" refers to the covalent or noncovalent joining of a polynucleotide, polypeptide, or antibody with a reporter molecule capable of producing a detectable or measurable signal.

"Microarray" is any arrangement of nucleic acids, amino acids, antibodies, etc., on a substrate. The substrate may be a solid support such as beads, glass, paper, nitrocellulose, nylon, or an appropriate membrane.

"Linkers" are short stretches of nucleotide sequence which may be added to a vector or an mddt to create restriction endonuclease sites to facilitate cloning. "Polylinkers" are engineered to incorporate multiple restriction enzyme sites and to provide for the use of enzymes which leave 5' or 3' overhangs (e.g., BamHI, EcoRI, and HindIII) and those which provide blunt ends (e.g., EcoRV, SnaBI, and StuI).

"Naturally occurring" refers to an endogenous polynucleotide or polypeptide that may be isolated from viruses or prokaryotic or eukaryotic cells.

"Nucleic acid sequence" refers to the specific order of nucleotides joined by phosphodiester bonds in a linear, polymeric arrangement. Depending on the number of nucleotides, the nucleic acid sequence can be considered an oligomer, oligonucleotide, or polynucleotide. The nucleic acid can be DNA, RNA, or any nucleic acid analog, such as PNA, may be of genomic or synthetic origin, may be either double-stranded or single-stranded, and can represent either the sense or antisense (complementary) strand.

"Oligomer" refers to a nucleic acid sequence of at least about 6 nucleotides and as many as about 60 nucleotides, preferably about 15 to 40 nucleotides, and most preferably between about 20 and 30 nucleotides, that may be used in hybridization or amplification technologies. Oligomers may be used as, e.g., primers for PCR, and are usually chemically synthesized.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to a DNA mimic in which nucleotide bases are attached to a pseudopeptide backbone to increase stability. PNAs, also designated antigene agents, can prevent gene expression by targeting complementary messenger RNA.

The phrases "percent identity" and "% identity", as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and Sharp, P.M. (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequence pairs.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to determine alignment between a known polynucleotide sequence and other sequences on a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2/. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in figures or Sequence Listings, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity", as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the hydrophobicity and acidity of the substituted residue, thus preserving the structure (and therefore function) of the folded polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalty

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for

example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in figures or Sequence Listings, may be used to describe a length over which percentage identity may be measured.

"Post-translational modification" of an MDDT may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu and the MDDT.

"Probe" refers to mddt or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the figures and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel et al.,1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of

Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

"Purified" refers to molecules, either polynucleotides or polypeptides that are isolated or separated from their natural environment and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other compounds with which they are naturally associated.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

"Regulatory element" refers to a nucleic acid sequence from nontranslated regions of a gene, and includes enhancers, promoters, introns, and 3' untranslated regions, which interact with host proteins to carry out or regulate transcription or translation.

"Reporter" molecules are chemical or biochemical moieties used for labeling a nucleic acid, an amino acid, or an antibody. They include radionuclides; enzymes; fluorescent, chemiluminescent,

or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

"Sample" is used in its broadest sense. Samples may contain nucleic or amino acids, antibodies, or other materials, and may be derived from any source (e.g., bodily fluids including, but not limited to, saliva, blood, and urine; chromosome(s), organelles, or membranes isolated from a cell; genomic DNA, RNA, or cDNA in solution or bound to a substrate; and cleared cells or tissues or blots or imprints from such cells or tissues).

"Specific binding" or "specifically binding" refers to the interaction between a protein or peptide and its agonist, antibody, antagonist, or other binding partner. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

"Substitution" refers to the replacement of at least one nucleotide or amino acid by a different nucleotide or amino acid.

"Substrate" refers to any suitable rigid or semi-rigid support including, e.g., membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles or capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular tissue or cell type under given conditions at a given time.

"Transformation" refers to a process by which exogenous DNA enters a recipient cell. Transformation may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the host cell being transformed.

"Transformants" include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as cells which transiently express inserted DNA or RNA.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid

introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, and plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 25% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 30%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95% or even at least 98% or greater sequence identity over a certain defined length. The variant may result in "conservative" amino acid changes which do not affect structural and/or chemical properties. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

In an alternative, variants of the polynucleotides of the present invention may be generated through recombinant methods. One possible method is a DNA shuffling technique such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of MDDT, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene

variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

In a particular embodiment, cDNA sequences derived from human tissues and cell lines were aligned based on nucleotide sequence identity and assembled into "consensus" or "template" sequences which are designated by the template identification numbers (template IDs) in column 2 of Table 1. The sequence identification numbers (SEQ ID NO:s) corresponding to the template IDs are shown in column 1. The template sequences have similarity to GenBank sequences, or "hits," as designated by the GI Numbers in column 3. The statistical probability of each GenBank hit is indicated by a probability score in column 4, and the functional annotation corresponding to each GenBank hit is listed in column 5.

The invention incorporates the nucleic acid sequences of these templates as disclosed in the Sequence Listing and the use of these sequences in the diagnosis and treatment of disease states characterized by defects in disease detection and treatment molecule molecules. The invention further utilizes these sequences in hybridization and amplification technologies, and in particular, in technologies which assess gene expression patterns correlated with specific cells or tissues and their responses in vivo or in vitro to pharmaceutical agents, toxins, and other treatments. In this manner, the sequences of the present invention are used to develop a transcript image for a particular cell or tissue.

Derivation of Nucleic Acid Sequences

cDNA was isolated from libraries constructed using RNA derived from normal and diseased human tissues and cell lines. The human tissues and cell lines used for cDNA library construction were selected from a broad range of sources to provide a diverse population of cDNAs representative of gene transcription throughout the human body. Descriptions of the human tissues and cell lines used for cDNA library construction are provided in the LIFESEQ database (Incyte Genomics, Inc. (Incyte), Palo Alto CA). Human tissues were broadly selected from, for example, cardiovascular, dermatologic, endocrine, gastrointestinal, hematopoietic/immune system, musculoskeletal, neural, reproductive, and urologic sources.

Cell lines used for cDNA library construction were derived from, for example, leukemic cells, teratocarcinomas, neuroepitheliomas, cervical carcinoma, lung fibroblasts, and endothelial cells. Such cell lines include, for example, THP-1, Jurkat, HUVEC, hNT2, WI38, HeLa, and other cell lines commonly used and available from public depositories (American Type Culture Collection, Manassas VA). Prior to mRNA isolation, cell lines were untreated, treated with a pharmaceutical agent such as 5'-aza-2'-deoxycytidine, treated with an activating agent such as lipopolysaccharide in the case of leukocytic cell lines, or, in the case of endothelial cell lines, subjected to shear stress.

Sequencing of the cDNAs

Methods for DNA sequencing are well known in the art. Conventional enzymatic methods employ the Klenow fragment of DNA polymerase I, SEQUENASE DNA polymerase (U.S. Biochemical Corporation, Cleveland OH), Taq polymerase (PE Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Inc. (Amersham Pharmacia Biotech), Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies Inc. (Life Technologies), Gaithersburg MD), to extend the nucleic acid sequence from an oligonucleotide primer annealed to the DNA template of interest. Methods have been developed for the use of both single-stranded and double-stranded templates. Chain termination reaction products may be electrophoresed on urea-polyacrylamide gels and detected either by autoradiography (for radioisotope-labeled nucleotides) or by fluorescence (for fluorophore-labeled nucleotides). Automated methods for mechanized reaction preparation, sequencing, and analysis using fluorescence detection methods have been developed. Machines used to prepare cDNAs for sequencing can include the MICROLAB 2200 liquid transfer system (Hamilton Company (Hamilton), Reno NV), Peltier thermal cycler (PTC200; MJ Research, Inc. (MJ Research), Watertown MA), and ABI CATALYST 800 thermal cycler (PE Biosystems). Sequencing can be carried out using, for example, the ABI 373 or 377 (PE Biosystems) or MEGABACE 1000 (Molecular Dynamics, Inc. (Molecular Dynamics), Sunnyvale CA) DNA sequencing systems, or other automated and manual sequencing systems well known in the art.

The nucleotide sequences of the Sequence Listing have been prepared by current, state-of-the-art, automated methods and, as such, may contain occasional sequencing errors or unidentified nucleotides. Such unidentified nucleotides are designated by an N. These infrequent unidentified bases do not represent a hindrance to practicing the invention for those skilled in the art. Several methods employing standard recombinant techniques may be used to correct errors and complete the missing sequence information. (See, e.g., those described in Ausubel, F.M. et al. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY; and Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY.)

Assembly of cDNA Sequences

Human polynucleotide sequences may be assembled using programs or algorithms well known in the art. Sequences to be assembled are related, wholly or in part, and may be derived from a single or many different transcripts. Assembly of the sequences can be performed using such programs as PHRAP (Phils Revised Assembly Program) and the GELVIEW fragment assembly system (GCG), or other methods known in the art.

Alternatively, cDNA sequences are used as "component" sequences that are assembled into "template" or "consensus" sequences as follows. Sequence chromatograms are processed, verified, and quality scores are obtained using PHRED. Raw sequences are edited using an editing pathway known as Block 1 (See, e.g., the LIFESEQ Assembled User Guide, Incyte Genomics, Palo Alto, CA). A series of BLAST comparisons is performed and low-information segments and repetitive elements (e.g., dinucleotide repeats, Alu repeats, etc.) are replaced by "n's", or masked, to prevent spurious matches. Mitochondrial and ribosomal RNA sequences are also removed. The processed sequences are then loaded into a relational database management system (RDMS) which assigns edited sequences to existing templates, if available. When additional sequences are added into the RDMS, a process is initiated which modifies existing templates or creates new templates from works in progress (i.e., nonfinal assembled sequences) containing queued sequences or the sequences themselves. After the new sequences have been assigned to templates, the templates can be merged into bins. If multiple templates exist in one bin, the bin can be split and the templates reannotated.

Once gene bins have been generated based upon sequence alignments, bins are "clone joined" based upon clone information. Clone joining occurs when the 5' sequence of one clone is present in one bin and the 3' sequence from the same clone is present in a different bin, indicating that the two bins should be merged into a single bin. Only bins which share at least two different clones are merged.

A resultant template sequence may contain either a partial or a full length open reading frame, or all or part of a genetic regulatory element. This variation is due in part to the fact that the full length cDNAs of many genes are several hundred, and sometimes several thousand, bases in

length. With current technology, cDNAs comprising the coding regions of large genes cannot be cloned because of vector limitations, incomplete reverse transcription of the mRNA, or incomplete "second strand" synthesis. Template sequences may be extended to include additional contiguous sequences derived from the parent RNA transcript using a variety of methods known to those of skill in the art. Extension may thus be used to achieve the full length coding sequence of a gene.

Analysis of the cDNA Sequences

The cDNA sequences are analyzed using a variety of programs and algorithms which are well known in the art. (See, e.g., Ausubel, 1997, supra, Chapter 7.7; Meyers, R.A. (Ed.) (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853; and Table 6.) These analyses comprise both reading frame determinations, e.g., based on triplet codon periodicity for particular organisms (Fickett, J.W. (1982) Nucleic Acids Res. 10:5303-5318); analyses of potential start and stop codons; and homology searches.

Computer programs known to those of skill in the art for performing computer-assisted searches for amino acid and nucleic acid sequence similarity, include, for example, Basic Local Alignment Search Tool (BLAST; Altschul, S.F. (1993) J. Mol. Evol. 36:290-300; Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410). BLAST is especially useful in determining exact matches and comparing two sequence fragments of arbitrary but equal lengths, whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user (Karlin, S. et al. (1988) Proc. Natl. Acad. Sci. USA 85:841-845). Using an appropriate search tool (e.g., BLAST or HMM), GenBank, SwissProt, BLOCKS, PFAM and other databases may be searched for sequences containing regions of homology to a query mddt or MDDT of the present invention.

Other approaches to the identification, assembly, storage, and display of nucleotide and polypeptide sequences are provided in "Relational Database for Storing Biomolecule Information," U.S.S.N. 08/947,845, filed October 9, 1997; "Project-Based Full-Length Biomolecular Sequence Database," U.S.S.N. 08/811,758, filed March 6, 1997; and "Relational Database and System for Storing Information Relating to Biomolecular Sequences," U.S.S.N. 09/034,807, filed March 4, 1998, all of which are incorporated by reference herein in their entirety.

Protein hierarchies can be assigned to the putative encoded polypeptide based on, e.g., motif, BLAST, or biological analysis. Methods for assigning these hierarchies are described, for example, in "Database System Employing Protein Function Hierarchies for Viewing Biomolecular Sequence Data," U.S.S.N. 08/812,290, filed March 6, 1997, incorporated herein by reference.

Human Disease Detection and Treatment Molecule Sequences

The mddt of the present invention may be used for a variety of diagnostic and therapeutic

purposes. For example, an mddt may be used to diagnose a particular condition, disease, or disorder associated with disease detection and treatment molecules. Such conditions, diseases, and disorders include, but are not limited to, a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an autoimmune/inflammatory disorder, such as actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma. The mddt can be used to detect the presence of, or to quantify the amount of, an mddt-related polynucleotide in a sample. This information is then compared to information obtained from appropriate reference samples, and a diagnosis is established. Alternatively, a polynucleotide complementary to a given mddt can inhibit or inactivate a therapeutically relevant gene related to the mddt.

Analysis of mddt Expression Patterns

The expression of mddt may be routinely assessed by hybridization-based methods to determine, for example, the tissue-specificity, disease-specificity, or developmental stage-specificity of mddt expression. For example, the level of expression of mddt may be compared among different cell types or tissues, among diseased and normal cell types or tissues, among cell types or tissues at different developmental stages, or among cell types or tissues undergoing various treatments. This type of analysis is useful, for example, to assess the relative levels of mddt expression in fully or

partially differentiated cells or tissues, to determine if changes in mddt expression levels are correlated with the development or progression of specific disease states, and to assess the response of a cell or tissue to a specific therapy, for example, in pharmacological or toxicological studies. Methods for the analysis of mddt expression are based on hybridization and amplification technologies and include membrane-based procedures such as northern blot analysis, high-throughput procedures that utilize, for example, microarrays, and PCR-based procedures.

Hybridization and Genetic Analysis

The mddt, their fragments, or complementary sequences, may be used to identify the presence of and/or to determine the degree of similarity between two (or more) nucleic acid sequences. The mddt may be hybridized to naturally occurring or recombinant nucleic acid sequences under appropriately selected temperatures and salt concentrations. Hybridization with a probe based on the nucleic acid sequence of at least one of the mddt allows for the detection of nucleic acid sequences, including genomic sequences, which are identical or related to the mddt of the Sequence Listing. Probes may be selected from non-conserved or unique regions of at least one of the polynucleotides of SEQ ID NO:1-25 and tested for their ability to identify or amplify the target nucleic acid sequence using standard protocols.

Polynucleotide sequences that are capable of hybridizing, in particular, to those shown in SEQ ID NO:1-25 and fragments thereof, can be identified using various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions are discussed in "Definitions."

A probe for use in Southern or northern hybridization may be derived from a fragment of an mddt sequence, or its complement, that is up to several hundred nucleotides in length and is either single-stranded or double-stranded. Such probes may be hybridized in solution to biological materials such as plasmids, bacterial, yeast, or human artificial chromosomes, cleared or sectioned tissues, or to artificial substrates containing mddt. Microarrays are particularly suitable for identifying the presence of and detecting the level of expression for multiple genes of interest by examining gene expression correlated with, e.g., various stages of development, treatment with a drug or compound, or disease progression. An array analogous to a dot or slot blot may be used to arrange and link polynucleotides to the surface of a substrate using one or more of the following: mechanical (vacuum), chemical, thermal, or UV bonding procedures. Such an array may contain any number of mddt and may be produced by hand or by using available devices, materials, and machines.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-

2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

Probes may be labeled by either PCR or enzymatic techniques using a variety of commercially available reporter molecules. For example, commercial kits are available for radioactive and chemiluminescent labeling (Amersham Pharmacia Biotech) and for alkaline phosphatase labeling (Life Technologies). Alternatively, mddt may be cloned into commercially available vectors for the production of RNA probes. Such probes may be transcribed in the presence of at least one labeled nucleotide (e.g., ^{32}P-ATP, Amersham Pharmacia Biotech).

Additionally the polynucleotides of SEQ ID NO:1-25 or suitable fragments thereof can be used to isolate full length cDNA sequences utilizing hybridization and/or amplification procedures well known in the art, e.g., cDNA library screening, PCR amplification, etc. The molecular cloning of such full length cDNA sequences may employ the method of cDNA library screening with probes using the hybridization, stringency, washing, and probing strategies described above and in Ausubel, supra, Chapters 3, 5, and 6. These procedures may also be employed with genomic libraries to isolate genomic sequences of mddt in order to analyze, e.g., regulatory elements.

Genetic Mapping

Gene identification and mapping are important in the investigation and treatment of almost all conditions, diseases, and disorders. Cancer, cardiovascular disease, Alzheimer's disease, arthritis, diabetes, and mental illnesses are of particular interest. Each of these conditions is more complex than the single gene defects of sickle cell anemia or cystic fibrosis, with select groups of genes being predictive of predisposition for a particular condition, disease, or disorder. For example, cardiovascular disease may result from malfunctioning receptor molecules that fail to clear cholesterol from the bloodstream, and diabetes may result when a particular individual's immune system is activated by an infection and attacks the insulin-producing cells of the pancreas. In some studies, Alzheimer's disease has been linked to a gene on chromosome 21; other studies predict a different gene and location. Mapping of disease genes is a complex and reiterative process and generally proceeds from genetic linkage analysis to physical mapping.

As a condition is noted among members of a family, a genetic linkage map traces parts of chromosomes that are inherited in the same pattern as the condition. Statistics link the inheritance of particular conditions to particular regions of chromosomes, as defined by RFLP or other markers. (See, for example, Lander, E. S. and Botstein, D. (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.) Occasionally, genetic markers and their locations are known from previous studies. More often, however, the markers are simply stretches of DNA that differ among individuals. Examples of genetic linkage maps can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site.

In another embodiment of the invention, mddt sequences may be used to generate

hybridization probes useful in chromosomal mapping of naturally occurring genomic sequences. Either coding or noncoding sequences of mddt may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of an mddt coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Meyers, supra, pp. 965-968.) Correlation between the location of mddt on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The mddt sequences may also be used to detect polymorphisms that are genetically linked to the inheritance of a particular condition, disease, or disorder.

In situ hybridization of chromosomal preparations and genetic mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending existing genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of the corresponding human chromosome is not known. These new marker sequences can be mapped to human chromosomes and may provide valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once a disease or syndrome has been crudely correlated by genetic linkage with a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequences of the subject invention may also be used to detect differences in chromosomal architecture due to translocation, inversion, etc., among normal, carrier, or affected individuals.

Once a disease-associated gene is mapped to a chromosomal region, the gene must be cloned in order to identify mutations or other alterations (e.g., translocations or inversions) that may be correlated with disease. This process requires a physical map of the chromosomal region containing the disease-gene of interest along with associated markers. A physical map is necessary for determining the nucleotide sequence of and order of marker genes on a particular chromosomal region. Physical mapping techniques are well known in the art and require the generation of overlapping sets of cloned DNA fragments from a particular organelle, chromosome, or genome. These clones are analyzed to reconstruct and catalog their order. Once the position of a marker is

determined, the DNA from that region is obtained by consulting the catalog and selecting clones from that region. The gene of interest is located through positional cloning techniques using hybridization or similar methods.

Diagnostic Uses

The mddt of the present invention may be used to design probes useful in diagnostic assays. Such assays, well known to those skilled in the art, may be used to detect or confirm conditions, disorders, or diseases associated with abnormal levels of mddt expression. Labeled probes developed from mddt sequences are added to a sample under hybridizing conditions of desired stringency. In some instances, mddt, or fragments or oligonucleotides derived from mddt, may be used as primers in amplification steps prior to hybridization. The amount of hybridization complex formed is quantified and compared with standards for that cell or tissue. If mddt expression varies significantly from the standard, the assay indicates the presence of the condition, disorder, or disease. Qualitative or quantitative diagnostic methods may include northern, dot blot, or other membrane or dip-stick based technologies or multiple-sample format technologies such as PCR, enzyme-linked immunosorbent assay (ELISA)-like, pin, or chip-based assays.

The probes described above may also be used to monitor the progress of conditions, disorders, or diseases associated with abnormal levels of mddt expression, or to evaluate the efficacy of a particular therapeutic treatment. The candidate probe may be identified from the mddt that are specific to a given human tissue and have not been observed in GenBank or other genome databases. Such a probe may be used in animal studies, preclinical tests, clinical trials, or in monitoring the treatment of an individual patient. In a typical process, standard expression is established by methods well known in the art for use as a basis of comparison, samples from patients affected by the disorder or disease are combined with the probe to evaluate any deviation from the standard profile, and a therapeutic agent is administered and effects are monitored to generate a treatment profile. Efficacy is evaluated by determining whether the expression progresses toward or returns to the standard normal pattern. Treatment profiles may be generated over a period of several days or several months. Statistical methods well known to those skilled in the art may be use to determine the significance of such therapeutic agents.

The polynucleotides are also useful for identifying individuals from minute biological samples, for example, by matching the RFLP pattern of a sample's DNA to that of an individual's DNA. The polynucleotides of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, an individual can be identified through a unique set of DNA sequences. Once a unique ID database is established for an individual, positive identification of that

individual can be made from extremely small tissue samples.

In a particular aspect, oligonucleotide primers derived from the mddt of the invention may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from mddt are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequences of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

DNA-based identification techniques are critical in forensic technology. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using, e.g., PCR, to identify individuals. (See, e.g., Erlich, H. (1992) PCR Technology, Freeman and Co., New York, NY). Similarly, polynucleotides of the present invention can be used as polymorphic markers.

There is also a need for reagents capable of identifying the source of a particular tissue. Appropriate reagents can comprise, for example, DNA probes or primers prepared from the sequences of the present invention that are specific for particular tissues. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

The polynucleotides of the present invention can also be used as molecular weight markers on nucleic acid gels or Southern blots, as diagnostic probes for the presence of a specific mRNA in a particular cell type, in the creation of subtracted cDNA libraries which aid in the discovery of novel polynucleotides, in selection and synthesis of oligomers for attachment to an array or other support, and as an antigen to elicit an immune response.

Disease Model Systems Using mddt

The mddt of the invention or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

The mddt of the invention may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

The mddt of the invention can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of mddt is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress mddt, resulting, e.g., in the secretion of MDDT in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

Screening Assays

MDDT encoded by polynucleotides of the present invention may be used to screen for molecules that bind to or are bound by the encoded polypeptides. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit (antagonist), or decrease activity of the polypeptide or the bound molecule. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a ligand or fragment thereof, a natural substrate, or a structural or functional mimetic. (See, Coligan et al., (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or to at least a fragment of the receptor, e.g., the active site. In either case, the molecule can be rationally designed using known techniques. Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing the polypeptide or cell membrane fractions which contain the expressed polypeptide are then contacted with a test compound and binding, stimulation, or inhibition of activity of either the polypeptide or the molecule is analyzed.

An assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. Alternatively, the assay may assess binding in the presence of a labeled competitor.

Additionally, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay using, e.g., a monoclonal or polyclonal antibody, can measure polypeptide level in a sample. The antibody can measure polypeptide level by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of the above assays can be used in a diagnostic or prognostic context. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Transcript Imaging and Toxicological Testing

Another embodiment relates to the use of mddt to develop a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput

format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity pertaining to disease detection and treatment molecules.

Transcript images which profile mddt expression may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect mddt expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile mddt expression may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E. F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and Anderson, N. L. (2000) Toxicol. Lett. 112-113:467-71, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of MDDT encoded by polynucleotides of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a

proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for MDDT to quantify the levels of MDDT expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-11; Mendoze, L. G. et al. (1999) Biotechniques 27:778-88). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N. L. and Seilhamer, J. (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated

biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the MDDT encoded by polynucleotides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the MDDT encoded by polynucleotides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Transcript images may be used to profile mddt expression in distinct tissue types. This process can be used to determine disease detection and treatment molecule activity in a particular tissue type relative to this activity in a different tissue type. Transcript images may be used to generate a profile of mddt expression characteristic of diseased tissue. Transcript images of tissues before and after treatment may be used for diagnostic purposes, to monitor the progression of disease, and to monitor the efficacy of drug treatments for diseases which affect the activity of disease detection and treatment molecules.

Transcript images of cell lines can be used to assess disease detection and treatment molecule activity and/or to identify cell lines that lack or misregulate this activity. Such cell lines may then be treated with pharmaceutical agents, and a transcript image following treatment may indicate the efficacy of these agents in restoring desired levels of this activity. A similar approach may be used to assess the toxicity of pharmaceutical agents as reflected by undesirable changes in disease detection and treatment molecule activity. Candidate pharmaceutical agents may be evaluated by comparing their associated transcript images with those of pharmaceutical agents of known effectiveness.

Antisense Molecules

The polynucleotides of the present invention are useful in antisense technology. Antisense technology or therapy relies on the modulation of expression of a target protein through the specific binding of an antisense sequence to a target sequence encoding the target protein or directing its expression. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ; Alama, A. et al. (1997) Pharmacol. Res. 36(3):171-178; Crooke, S.T. (1997) Adv. Pharmacol. 40:1-49; Sharma, H.W. and R. Narayanan (1995) Bioessays 17(12):1055-1063; and Lavrosky, Y. et al. (1997) Biochem. Mol. Med. 62(1):11-22.) An antisense sequence is a polynucleotide sequence

capable of specifically hybridizing to at least a portion of the target sequence. Antisense sequences bind to cellular mRNA and/or genomic DNA, affecting translation and/or transcription. Antisense sequences can be DNA, RNA, or nucleic acid mimics and analogs. (See, e.g., Rossi, J.J. et al. (1991) Antisense Res. Dev. 1(3):285-288; Lee, R. et al. (1998) Biochemistry 37(3):900-1010; Pardridge, W.M. et al. (1995) Proc. Natl. Acad. Sci. USA 92(12):5592-5596; and Nielsen, P. E. and Haaima, G. (1997) Chem. Soc. Rev. 96:73-78.) Typically, the binding which results in modulation of expression occurs through hybridization or binding of complementary base pairs. Antisense sequences can also bind to DNA duplexes through specific interactions in the major groove of the double helix.

The polynucleotides of the present invention and fragments thereof can be used as antisense sequences to modify the expression of the polypeptide encoded by mddt. The antisense sequences can be produced ex vivo, such as by using any of the ABI nucleic acid synthesizer series (PE Biosystems) or other automated systems known in the art. Antisense sequences can also be produced biologically, such as by transforming an appropriate host cell with an expression vector containing the sequence of interest. (See, e.g., Agrawal, supra.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E., et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J., et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

Expression

In order to express a biologically active MDDT, the nucleotide sequences encoding MDDT or fragments thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding MDDT and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, supra, Chapters 4, 8,

16, and 17; and Ausubel, supra, Chapters 9, 10, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding MDDT. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal (mammalian) cell systems. (See, e.g., Sambrook, supra; Ausubel, 1995, supra, Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al., (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

For long term production of recombinant proteins in mammalian systems, stable expression of MDDT in cell lines is preferred. For example, sequences encoding MDDT can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Any number of selection systems may be used to recover transformed cell lines. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.; Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14; Hartman, S.C. and R.C.Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051; Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Therapeutic Uses of mddt

The mddt of the invention may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et

al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and Somia, N. (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in mddt expression or regulation causes disease, the expression of mddt from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in mddt are treated by constructing mammalian expression vectors comprising mddt and introducing these vectors by mechanical means into mddt-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and Anderson, W.F. (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and Récipon, H. (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of mddt include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). The mddt of the invention may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and Bujard, H. (1992) Proc. Natl. Acad. Sci. U.S.A. 89:5547-5551; Gossen, M. et al., (1995) Science 268:1766-1769; Rossi, F.M.V. and Blau, H.M. (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding MDDT from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and Eb, A.J. (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to mddt expression are treated by constructing a retrovirus vector consisting of (i) mddt under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. U.S.A. 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and Miller, A.D. (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. U.S.A. 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver mddt to cells which have one or more genetic abnormalities with respect to the expression of mddt. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and Somia, N. (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver mddt to target cells which have one or more genetic abnormalities with respect to the expression of mddt. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing mddt to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res.169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W. F. et al. 1999 J. Virol. 73:519-532 and Xu, H. et al., (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver mddt to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and Li, K-J. (1998) Curr. Opin. Biotech. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting mddt into the alphavirus genome in place of the capsid-coding region results in the production of a large number of mddt RNAs and the synthesis of high levels of MDDT in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of mddt into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Antibodies

Anti-MDDT antibodies may be used to analyze protein expression levels. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, and Fab fragments. For descriptions of and protocols of antibody technologies, see, e.g., Pound J.D. (1998) Immunochemical Protocols, Humana Press, Totowa, NJ.

The amino acid sequence encoded by the mddt of the Sequence Listing may be analyzed by appropriate software (e.g., LASERGENE NAVIGATOR software, DNASTAR) to determine regions of high immunogenicity. The optimal sequences for immunization are selected from the C-terminus, the N-terminus, and those intervening, hydrophilic regions of the polypeptide which are likely to be exposed to the external environment when the polypeptide is in its natural conformation. Analysis used to select appropriate epitopes is also described by Ausubel (1997, supra, Chapter 11.7). Peptides used for antibody induction do not need to have biological activity; however, they must be antigenic. Peptides used to induce specific antibodies may have an amino acid sequence consisting of at five amino acids, preferably at least 10 amino acids, and most preferably 15 amino acids. A peptide which mimics an antigenic fragment of the natural polypeptide may be fused with another protein such as keyhole limpet cyanin (KLH; Sigma, St. Louis MO) for antibody production. A peptide encompassing an antigenic region may be expressed from an mddt, synthesized as described above, or purified from human cells.

Procedures well known in the art may be used for the production of antibodies. Various hosts including mice, goats, and rabbits, may be immunized by injection with a peptide. Depending on the host species, various adjuvants may be used to increase immunological response.

In one procedure, peptides about 15 residues in length may be synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using fmoc-chemistry and coupled to KLH (Sigma) by reaction with M-maleimidobenzoyl-N-hydroxysuccinimide ester (Ausubel, 1995, supra). Rabbits are immunized with the peptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity by binding the peptide to plastic, blocking with 1% bovine serum albumin (BSA), reacting with rabbit antisera, washing, and reacting with radioiodinated goat anti-rabbit IgG. Antisera with antipeptide activity are tested for anti-MDDT activity using protocols well known in the art, including ELISA, radioimmunoassay (RIA), and immunoblotting.

In another procedure, isolated and purified peptide may be used to immunize mice (about 100 µg of peptide) or rabbits (about 1 mg of peptide). Subsequently, the peptide is radioiodinated and used to screen the immunized animals' B-lymphocytes for production of antipeptide antibodies. Positive cells are then used to produce hybridomas using standard techniques. About 20 mg of peptide is sufficient for labeling and screening several thousand clones. Hybridomas of interest are detected by screening with radioiodinated peptide to identify those fusions producing peptide-specific monoclonal antibody. In a typical protocol, wells of a multi-well plate (FAST, Becton-Dickinson,

Palo Alto, CA) are coated with affinity-purified, specific rabbit-anti-mouse (or suitable anti-species IgG) antibodies at 10 mg/ml. The coated wells are blocked with 1% BSA and washed and exposed to supernatants from hybridomas. After incubation, the wells are exposed to radiolabeled peptide at 1 mg/ml.

Clones producing antibodies bind a quantity of labeled peptide that is detectable above background. Such clones are expanded and subjected to 2 cycles of cloning. Cloned hybridomas are injected into pristane-treated mice to produce ascites, and monoclonal antibody is purified from the ascitic fluid by affinity chromatography on protein A (Amersham Pharmacia Biotech). Several procedures for the production of monoclonal antibodies, including in vitro production, are described in Pound (supra). Monoclonal antibodies with antipeptide activity are tested for anti-MDDT activity using protocols well known in the art, including ELISA, RIA, and immunoblotting.

Antibody fragments containing specific binding sites for an epitope may also be generated. For example, such fragments include, but are not limited to, the F(ab')2 fragments produced by pepsin digestion of the antibody molecule, and the Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, construction of Fab expression libraries in filamentous bacteriophage allows rapid and easy identification of monoclonal fragments with desired specificity (Pound, supra, Chaps. 45-47). Antibodies generated against polypeptide encoded by mddt can be used to purify and characterize full-length MDDT protein and its activity, binding partners, etc.

Assays Using Antibodies

Anti-MDDT antibodies may be used in assays to quantify the amount of MDDT found in a particular human cell. Such assays include methods utilizing the antibody and a label to detect expression level under normal or disease conditions. The peptides and antibodies of the invention may be used with or without modification or labeled by joining them, either covalently or noncovalently, with a reporter molecule.

Protocols for detecting and measuring protein expression using either polyclonal or monoclonal antibodies are well known in the art. Examples include ELISA, RIA, and fluorescent activated cell sorting (FACS). Such immunoassays typically involve the formation of complexes between the MDDT and its specific antibody and the measurement of such complexes. These and other assays are described in Pound (supra).

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/156,565 and U.S. Ser. No. 60/168,197 are hereby expressly incorporated

by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from CLONTECH Laboratories, Inc. (Palo Alto CA) or isolated from various tissues. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In most cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega Corporation (Promega), Madison WI), OLIGOTEX latex particles (QIAGEN, Inc. (QIAGEN), Valencia CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Inc., Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene Cloning Systems, Inc. (Stratagene), La Jolla CA) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, Chapters 5.1 through 6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), pSPORT1 plasmid (Life Technologies), or pINCY (Incyte). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: the Magic or WIZARD Minipreps DNA purification system (Promega); the AGTC Miniprep purification kit (Edge BioSystems, Gaithersburg MD); and the QIAWELL 8, QIAWELL 8 Plus, and QIAWELL 8 Ultra

plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit (QIAGEN). Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format. (Rao, V.B. (1994) Anal. Biochem. 216:1-14.) Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Inc. (Molecular Probes), Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 thermal cycler (PE Biosystems) or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific Corp., Sunnyvale CA) or the MICROLAB 2200 liquid transfer system (Hamilton). cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (PE Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, Chapter 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

IV. Assembly and Analysis of Sequences

Component sequences from chromatograms were subject to PHRED analysis and assigned a quality score. The sequences having at least a required quality score were subject to various pre-processing editing pathways to eliminate, e.g., low quality 3' ends, vector and linker sequences, polyA tails, Alu repeats, mitochondrial and ribosomal sequences, bacterial contamination sequences, and sequences smaller than 50 base pairs. In particular, low-information sequences and repetitive elements (e.g., dinucleotide repeats, Alu repeats, etc.) were replaced by "n's", or masked, to prevent spurious matches.

Processed sequences were then subject to assembly procedures in which the sequences were assigned to gene bins (bins). Each sequence could only belong to one bin. Sequences in each gene bin were assembled to produce consensus sequences (templates). Subsequent new sequences were

added to existing bins using BLASTn (v.1.4 WashU) and CROSSMATCH. Candidate pairs were identified as all BLAST hits having a quality score greater than or equal to 150. Alignments of at least 82% local identity were accepted into the bin. The component sequences from each bin were assembled using a version of PHRAP. Bins with several overlapping component sequences were assembled using DEEP PHRAP. The orientation (sense or antisense) of each assembled template was determined based on the number and orientation of its component sequences. Template sequences as disclosed in the sequence listing correspond to sense strand sequences (the "forward" reading frames), to the best determination. The complementary (antisense) strands are inherently disclosed herein. The component sequences which were used to assemble each template consensus sequence are listed in Tables 4A and 4B , along with their positions along the template nucleotide sequences.

Bins were compared against each other and those having local similarity of at least 82% were combined and reassembled. Reassembled bins having templates of insufficient overlap (less than 95% local identity) were re-split. Assembled templates were also subject to analysis by STITCHER/EXON MAPPER algorithms which analyze the probabilities of the presence of splice variants, alternatively spliced exons, splice junctions, differential expression of alternative spliced genes across tissue types or disease states, etc. These resulting bins were subject to several rounds of the above assembly procedures.

Once gene bins were generated based upon sequence alignments, bins were clone joined based upon clone information. If the 5' sequence of one clone was present in one bin and the 3' sequence from the same clone was present in a different bin, it was likely that the two bins actually belonged together in a single bin. The resulting combined bins underwent assembly procedures to regenerate the consensus sequences.

The final assembled templates were subsequently annotated using the following procedure. Template sequences were analyzed using BLASTn (v2.0, NCBI) versus gbpri (GenBank version 118). "Hits" were defined as an exact match having from 95% local identity over 200 base pairs through 100% local identity over 100 base pairs, or a homolog match having an E-value, i.e. a probability score, of $\leq 1 \times 10^{-8}$. The hits were subject to frameshift FASTx versus GENPEPT (GenBank version 118). (See Table 6). In this analysis, a homolog match was defined as having an E-value of $\leq 1 \times 10^{-8}$. The assembly method used above was described in "System and Methods for Analyzing Biomolecular Sequences," U.S.S.N. 09/276,534, filed March 25, 1999, and the LIFESEQ Gold user manual (Incyte) both incorporated by reference herein.

Following assembly, template sequences were subjected to motif, BLAST, and functional analyses, and categorized in protein hierarchies using methods described in, e.g., "Database System Employing Protein Function Hierarchies for Viewing Biomolecular Sequence Data," U.S.S.N. 08/812,290, filed March 6, 1997; "Relational Database for Storing Biomolecule Information," U.S.S.N. 08/947,845, filed October 9, 1997; "Project-Based Full-Length Biomolecular Sequence

Database," U.S.S.N. 08/811,758, filed March 6, 1997; and "Relational Database and System for Storing Information Relating to Biomolecular Sequences," U.S.S.N. 09/034,807, filed March 4, 1998, all of which are incorporated by reference herein.

The template sequences were further analyzed by translating each template in all three forward reading frames and searching each translation against the Pfam database of hidden Markov model-based protein families and domains using the HMMER software package (available to the public from Washington University School of Medicine, St. Louis MO). Regions of templates which, when translated, contain similarity to Pfam consensus sequences are reported in Table 2, along with descriptions of Pfam protein domains and families. Only those Pfam hits with an E-value of $\leq 1 \times 10^{-3}$ are reported. (See also World Wide Web site http://pfam.wustl.edu/ for detailed descriptions of Pfam protein domains and families.)

Additionally, the template sequences were translated in all three forward reading frames, and each translation was searched against hidden Markov models for signal peptide and transmembrane domains using the HMMER software package. Construction of hidden Markov models and their usage in sequence analysis has been described. (See, for example, Eddy, S.R. (1996) Curr. Opin. Str. Biol. 6:361-365.) Regions of templates which, when translated, contain similarity to signal peptide or transmembrane domain consensus sequences are reported in Table 3. Only those signal peptide or transmembrane hits with a cutoff score of 11 bits or greater are reported. A cutoff score of 11 bits or greater corresponds to at least about 91-94% true-positives in signal peptide prediction, and at least about 75% true-positives in transmembrane domain prediction.

The results of HMMER analysis as reported in Tables 2 and 3 may support the results of BLAST analysis as reported in Table 1 or may suggest alternative or additional properties of template-encoded polypeptides not previously uncovered by BLAST or other analyses.

Template sequences are further analyzed using the bioinformatics tools listed in Table 6, or using sequence analysis software known in the art such as MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Template sequences may be further queried against public databases such as the GenBank rodent, mammalian, vertebrate, prokaryote, and eukaryote databases.

V. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is

much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

VI. Tissue Distribution Profiling

A tissue distribution profile is determined for each template by compiling the cDNA library tissue classifications of its component cDNA sequences. Each component sequence, is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. Template sequences, component sequences, and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

Table 5 shows the tissue distribution profile for the templates of the invention. For each template, the three most frequently observed tissue categories are shown in column 3, along with the percentage of component sequences belonging to each category. Only tissue categories with percentage values of $\geq 10\%$ are shown. A tissue distribution of "widely distributed" in column 3 indicates percentage values of $<10\%$ in all tissue categories.

VII. Transcript Image Analysis

Transcript images are generated as described in Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, incorporated herein by reference.

VIII. Extension of Polynucleotide Sequences and Isolation of a Full-length cDNA

Oligonucleotide primers designed using an mddt of the Sequence Listing are used to extend the nucleic acid sequence. One primer is synthesized to initiate 5' extension of the template, and the other primer, to initiate 3' extension of the template. The initial primers may be designed using OLIGO 4.06 software (National Biosciences, Inc. (National Biosciences), Plymouth MN), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations are avoided. Selected human cDNA libraries are used to extend the sequence. If more than one extension is necessary or desired, additional or nested sets of primers are designed.

High fidelity amplification is obtained by PCR using methods well known in the art. PCR is performed in 96-well plates using the PTC-200 thermal cycler (MJ Research). The reaction mix contains DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and ß-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ are as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well is determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v); Molecular Probes) dissolved in 1X Tris-EDTA (TE) and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Incorporated (Corning), Corning NY), allowing the DNA to bind to the reagent. The plate is scanned in a FLUOROSKAN II (Labsystems Oy) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture is analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions are successful in extending the sequence.

The extended nucleotides are desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides are separated on low concentration (0.6 to 0.8%) agarose gels, fragments are excised, and agar digested with AGAR ACE (Promega). Extended clones

are religated using T4 ligase (New England Biolabs, Inc., Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells are selected on antibiotic-containing media, individual colonies are picked and cultured overnight at 37°C in 384-well plates in LB/2x carbenicillin liquid media.

The cells are lysed, and DNA is amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA is quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries are reamplified using the same conditions as described above. Samples are diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems).

In like manner, the mddt is used to obtain regulatory sequences (promoters, introns, and enhancers) using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling of Probes and Southern Hybridization Analyses

Hybridization probes derived from the mddt of the Sequence Listing are employed for screening cDNAs, mRNAs, or genomic DNA. The labeling of probe nucleotides between 100 and 1000 nucleotides in length is specifically described, but essentially the same procedure may be used with larger cDNA fragments. Probe sequences are labeled at room temperature for 30 minutes using a T4 polynucleotide kinase, γ^{32}P-ATP, and 0.5X One-Phor-All Plus (Amersham Pharmacia Biotech) buffer and purified using a ProbeQuant G-50 Microcolumn (Amersham Pharmacia Biotech). The probe mixture is diluted to 10^7 dpm/μg/ml hybridization buffer and used in a typical membrane-based hybridization analysis.

The DNA is digested with a restriction endonuclease such as Eco RV and is electrophoresed through a 0.7% agarose gel. The DNA fragments are transferred from the agarose to nylon membrane (NYTRAN Plus, Schleicher & Schuell, Inc., Keene NH) using procedures specified by the manufacturer of the membrane. Prehybridization is carried out for three or more hours at 68°C, and hybridization is carried out overnight at 68°C. To remove non-specific signals, blots are sequentially washed at room temperature under increasingly stringent conditions, up to 0.1x saline sodium citrate (SSC) and 0.5% sodium dodecyl sulfate. After the blots are placed in a PHOSPHORIMAGER cassette (Molecular Dynamics) or are exposed to autoradiography film, hybridization patterns of standard and experimental lanes are compared. Essentially the same procedure is employed when

screening RNA.

X. Chromosome Mapping of mddt

The cDNA sequences which were used to assemble SEQ ID NO:1-25 are compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that match SEQ ID NO:1-25 are assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as PHRAP (Table 6). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon are used to determine if any of the clustered sequences have been previously mapped. Inclusion of a mapped sequence in a cluster will result in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location. The genetic map locations of SEQ ID NO:1-25 are described as ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters.

XI. Microarray Analysis

Probe Preparation from Tissue or Cell Samples

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and polyA$^+$ RNA is purified using the oligo (dT) cellulose method. Each polyA$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-dT primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng polyA$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control polyA$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA (W. Lei, unpublished). As quantitative controls, the control mRNAs at 0.002 ng, 0.02 ng, 0.2 ng, and 2 ng are diluted into reverse transcription reaction at ratios of 1:100,000, 1:10,000, 1:1000, 1:100 (w/w) to sample mRNA respectively. The control mRNAs are diluted into reverse transcription reaction at ratios of 1:3, 3:1, 1:10, 10:1, 1:25, 25:1 (w/w) to sample mRNA differential expression patterns. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Probes are purified using two

successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The probe is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester, PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford, MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 µl of probe mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The probe mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5x SSC in a corner of the chamber. The chamber containing the arrays is incubated for

about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the probe mix at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two probes from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood, MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The

software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XII. Complementary Nucleic Acids

Sequences complementary to the mddt are used to detect, decrease, or inhibit expression of the naturally occurring nucleotide. The use of oligonucleotides comprising from about 15 to 30 base pairs is typical in the art. However, smaller or larger sequence fragments can also be used. Appropriate oligonucleotides are designed from the mddt using OLIGO 4.06 software (National Biosciences) or other appropriate programs and are synthesized using methods standard in the art or ordered from a commercial supplier. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent transcription factor binding to the promoter sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding and processing of the transcript.

XIII. Expression of MDDT

Expression and purification of MDDT is accomplished using bacterial or virus-based expression systems. For expression of MDDT in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express MDDT upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of MDDT in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding MDDT by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See e.g., Engelhard, supra; and Sandig, supra.)

In most expression systems, MDDT is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from

MDDT at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak Company, Rochester NY). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, Chapters 10 and 16). Purified MDDT obtained by these methods can be used directly in the following activity assay.

XIV. Demonstration of MDDT Activity

MDDT, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled MDDT, washed, and any wells with labeled MDDT complex are assayed. Data obtained using different concentrations of MDDT are used to calculate values for the number, affinity, and association of MDDT with the candidate molecules.

Alternatively, molecules interacting with MDDT are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (CLONTECH).

MDDT may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XV. Functional Assays

MDDT function is assessed by expressing mddt at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen Corporation, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 µg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 µg of an additional plasmid containing sequences encoding a marker protein are co-transfected.

Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; CLONTECH), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of

the cells and other cellular properties.

FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of MDDT on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding MDDT and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Inc., Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding MDDT and other genes of interest can be analyzed by northern analysis or microarray techniques.

XVI. Production of Antibodies

MDDT substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e:g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the MDDT amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding peptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, Chapter 11.)

Typically, peptides 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using fmoc-chemistry and coupled to KLH (Sigma) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, supra.) Rabbits are immunized with the peptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity by, for example, binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG. Antisera with antipeptide activity are tested for anti-MDDT activity using protocols well known in the art, including ELISA, RIA, and immunoblotting.

XVII. Purification of Naturally Occurring MDDT Using Specific Antibodies

Naturally occurring or recombinant MDDT is substantially purified by immunoaffinity chromatography using antibodies specific for MDDT. An immunoaffinity column is constructed by covalently coupling anti-MDDT antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing MDDT are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of MDDT (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/MDDT binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and MDDT is collected.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

TABLE 1

SEQ ID NO:	Template ID	GI Number	Probability Score	Annotation
16	233624.11.dec	g1314560	9.00E-31	amyloid precursor protein-binding protein 1
7	246526.2.dec	g7542723	1.00E-168	DHHC1 protein (Homo sapiens)
5	345638.1.oct	g7406641	2.00E-90	EMeg32 protein (Mus musculus)
18	198840.3.dec	g643590	0	Human alternatively spliced mRNA for NACP (precursor of non-A beta component
4	197170.1.oct	g4389513	8.00E-45	Human homolog of Mus musculus wizL protein (AA 4-1561) (Homo sapiens)
11	040422.12.dec	g3341980	4.00E-66	huntingtin-interacting protein HYPA/FBP11
21	349415.4.dec	g533523	1.00E-159	MAGE-6 antigen (Homo sapiens)
22	474778.3.dec	g2077825	7.00E-62	MNK1 (Homo sapiens)
15	196774.3.dec	g6457278	1.00E-59	pre-B lymphocyte protein 3 (Homo sapiens)
14	059263.6.dec	g1694682	1.00E-116	Src-like adapter protein (Homo sapiens)
13	012432.5.dec	g1314316	2.00E-13	WD-40 motifs; up-regulated by thyroid hormone in tadpoles (Xenopus laevis)

TABLE 2

SEQ ID NO:	Template ID	Start	Stop	Frame	Pfam Hit	Pfam Description	E-value
1	348736.2.oct	265	450	forward 1	KRAB	PF01352 KRAB box	2.50E-07
2	025119.6.oct	179	367	forward 2	KRAB	PF01352 KRAB box	1.80E-28
3	474539.1.oct	2	280	forward 2	PH	PF00169 PH (pleckstrin homology) domain	2.10E-08
4	197170.1.oct	194	262	forward 2	zf-C2H2	PF00096 Zinc finger, C2H2 type	3.10E-08
5	345638.1.oct	248	640	forward 2	Acetyltransf	PF00583 Acetyltransferase (GNAT) family	0.00033
6	408784.1.dec	207	335	forward 3	UBA	UBA-domain	1.90E-06
7	246526.2.dec	570	764	forward 3	zf-DHHC	DHHC zinc finger domain	2.60E-34
8	200488.5.dec	89	619	forward 2	Peptidase_C15	Pyroglutamyl	3.30E-04
9	474878.1.dec	1003	1116	forward 1	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	1.50E-05
10	335916.2.dec	1053	1151	forward 3	ank	Ank repeat	1.10E-06
11	040422.12.dec	478	567	forward 1	WW_rsp5_WWP	WW domain	2.40E-12
12	977651.2.dec	718	924	forward 1	NifU-like	NifU-like domain	3.60E-30
13	012432.5.dec	280	396	forward 1	WD40	WD domain, G-beta repeat	7.00E-05
14	059263.6.dec	645	875	forward 3	SH2	Src homology domain	1.30E-33
15	196774.3.dec	695	949	forward 2	ig	Immunoglobulin	2.10E-09
16	233624.11.dec	345	656	forward 3	ThiF_family	ThiF family	4.00E-05
16	233624.11.dec	245	730	forward 2	ThiF_family	ThiF family	4.90E-04
17	228585.3.dec	927	1250	forward 3	PH	PH domain	1.50E-06
17	228585.3.dec	294	833	forward 3	RhoGEF	RhoGEF domain	7.00E-39
17	228585.3.dec	21	185	forward 3	SH3	Src homology domain	1.20E-08
18	198840.3.dec	137	502	forward 2	Synuclein	Synuclein	2.40E-72
19	082154.5.dec	50	340	forward 2	FCH	Fes/CIP4 homology domain	7.60E-05
20	368396.5.dec	3391	3555	forward 1	SH3	Src homology domain	2.40E-21
21	349415.4.dec	2408	3094	forward 2	MAGE	MAGE family	1.20E-134
22	474778.3.dec	297	542	forward 3	pkinase	Eukaryotic protein kinase domain	6.50E-13
23	330933.5.dec	209	604	forward 2	DAGKc	Diacylglycerol kinase catalytic domain (presumed)	4.80E-04
24	998036.2.dec	168	332	forward 3	SH3	Src homology domain	9.60E-20
24	998036.2.dec	956	1126	forward 2	SH3	Src homology domain	2.00E-17
25	999304.1.dec	78	218	forward 3	KRAB	KRAB box	2.30E-17

TABLE 3

SEQ ID NO:	Template ID	Start	Stop	Frame	Domain Type
5	345638.1.oct	1601	1657	forward 2	TM
5	345638.1.oct	243	296	forward 3	TM
7	246526.2.dec	366	419	forward 3	TM
7	246526.2.dec	738	812	forward 3	TM
7	246526.2.dec	738	797	forward 3	TM
7	246526.2.dec	375	452	forward 3	TM
7	246526.2.dec	855	911	forward 3	TM
7	246526.2.dec	849	923	forward 3	TM
7	246526.2.dec	861	938	forward 3	TM
7	246526.2.dec	735	797	forward 3	TM
7	246526.2.dec	855	908	forward 3	TM
7	246526.2.dec	2714	2797	forward 2	TM
9	474878.1.dec	1493	1561	forward 2	SP
9	474878.1.dec	126	194	forward 3	SP
9	474878.1.dec	852	902	forward 3	TM
9	474878.1.dec	2092	2163	forward 1	SP
9	474878.1.dec	1514	1573	forward 2	TM
10	335916.2.dec	579	638	forward 3	SP
10	335916.2.dec	555	638	forward 3	SP
10	335916.2.dec	1306	1389	forward 1	SP
11	040422.12.dec	865	933	forward 1	SP
11	040422.12.dec	945	1001	forward 3	SP
11	040422.12.dec	939	1007	forward 3	SP
11	040422.12.dec	939	1001	forward 3	TM
11	040422.12.dec	939	986	forward 3	SP
11	040422.12.dec	939	1001	forward 3	SP
11	040422.12.dec	945	1055	forward 3	SP
15	196774.3.dec	84	158	forward 3	SP
15	196774.3.dec	111	164	forward 3	TM
15	196774.3.dec	84	146	forward 3	SP
16	233624.11.dec	508	585	forward 1	SP
17	228585.3.dec	2343	2396	forward 3	TM
17	228585.3.dec	4942	4998	forward 1	SP
17	228585.3.dec	4975	5019	forward 1	SP
17	228585.3.dec	5218	5298	forward 1	SP
17	228585.3.dec	1633	1713	forward 1	SP
17	228585.3.dec	4417	4491	forward 1	SP
17	228585.3.dec	4942	5010	forward 1	SP
17	228585.3.dec	4942	5016	forward 1	SP
17	228585.3.dec	4975	5034	forward 1	SP
17	228585.3.dec	4942	5034	forward 1	SP
20	368396.5.dec	597	680	forward 3	SP
20	368396.5.dec	2585	2659	forward 2	SP
20	368396.5.dec	2585	2668	forward 2	SP
20	368396.5.dec	1051	1137	forward 1	SP
20	368396.5.dec	1051	1128	forward 1	SP
20	368396.5.dec	748	813	forward 1	SP
23	330933.5.dec	3492	3551	forward 3	TM
23	330933.5.dec	2174	2239	forward 2	TM
23	330933.5.dec	2627	2677	forward 2	TM

TABLE 3

SEQ ID NO:	Template ID	Start	Stop	Frame	Domain Type
23	330933.5.dec	2502	2552	forward 3	TM
23	330933.5.dec	2940	3026	forward 3	SP
23	330933.5.dec	2592	2651	forward 3	SP
23	330933.5.dec	2502	2549	forward 3	SP
23	330933.5.dec	2502	2567	forward 3	SP
23	330933.5.dec	2502	2555	forward 3	SP
23	330933.5.dec	2502	2561	forward 3	SP

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
1	348736.2.oct	899043R1	1	569
1	348736.2.oct	899043R6	29	569
1	348736.2.oct	g3597108	267	583
1	348736.2.oct	899072H1	270	569
1	348736.2.oct	899043H1	278	569
1	348736.2.oct	g2907503	297	473
1	348736.2.oct	g2903890	297	744
1	348736.2.oct	g2818919	297	736
1	348736.2.oct	g2904085	297	740
1	348736.2.oct	g2563340	297	595
1	348736.2.oct	g2817010	297	677
1	348736.2.oct	187645R6	10	105
1	348736.2.oct	187645R1	10	105
1	348736.2.oct	187645F1	10	106
1	348736.2.oct	187645H1	10	105
2	025119.6.oct	g2177786	304	651
2	025119.6.oct	g2434481	338	650
2	025119.6.oct	1568642H1	364	579
2	025119.6.oct	1572584H1	364	556
2	025119.6.oct	g3785307	410	673
2	025119.6.oct	g4136446	228	630
2	025119.6.oct	4828163H1	241	511
2	025119.6.oct	g2177785	256	622
2	025119.6.oct	g4223734	260	664
2	025119.6.oct	g2177771	270	625
2	025119.6.oct	g4087706	286	673
2	025119.6.oct	g1193161	291	672
2	025119.6.oct	g4223735	302	673
2	025119.6.oct	g2177772	304	631
2	025119.6.oct	3528954H1	1	225
2	025119.6.oct	3457794H1	1	240
2	025119.6.oct	g4124162	92	520
2	025119.6.oct	g2270206	115	551
2	025119.6.oct	1712170F6	140	628
2	025119.6.oct	1712170H1	140	358
2	025119.6.oct	g3076605	187	673
2	025119.6.oct	1616212H1	153	384
2	025119.6.oct	6110945H1	179	272
2	025119.6.oct	g3229162	182	656
2	025119.6.oct	3597144H1	184	464
2	025119.6.oct	4304325H1	150	369
2	025119.6.oct	5108047H1	152	383
3	474539.1.oct	g3039648	1	494
3	474539.1.oct	g4224114	1	444
3	474539.1.oct	g2354920	12	366
3	474539.1.oct	g2575314	34	417
3	474539.1.oct	g788735	42	278
3	474539.1.oct	g2753248	62	194
3	474539.1.oct	g1833029	145	334
3	474539.1.oct	5442680H1	258	429

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
4	197170.1.oct	2522574H1	378	630
4	197170.1.oct	6327035H1	470	573
4	197170.1.oct	1451166F1	494	779
4	197170.1.oct	1451166H1	494	767
4	197170.1.oct	1496387H1	512	724
4	197170.1.oct	g2537634	591	978
4	197170.1.oct	5324664H1	699	862
4	197170.1.oct	3274337H1	729	979
4	197170.1.oct	788329H1	878	990
4	197170.1.oct	1515193H1	881	1092
4	197170.1.oct	1515121H1	881	1081
4	197170.1.oct	1728752H1	898	1080
4	197170.1.oct	4671552H1	945	1197
4	197170.1.oct	g3931972	967	1429
4	197170.1.oct	g3428451	969	1392
4	197170.1.oct	g4267134	971	1429
4	197170.1.oct	g4534659	974	1392
4	197170.1.oct	g4111935	987	1392
4	197170.1.oct	g3919391	994	1429
4	197170.1.oct	g3419001	995	1392
4	197170.1.oct	g3896452	999	1392
4	197170.1.oct	g4300782	1017	1392
4	197170.1.oct	g3988440	1017	1429
4	197170.1.oct	g2056736	1021	1392
4	197170.1.oct	3234275H1	1046	1305
4	197170.1.oct	5163595H1	1097	1328
4	197170.1.oct	g4330857	1133	1392
4	197170.1.oct	g4194622	1168	1392
4	197170.1.oct	g3931900	1174	1429
4	197170.1.oct	g3049130	1213	1317
4	197170.1.oct	g3096022	1233	1393
4	197170.1.oct	g3888959	1277	1392
4	197170.1.oct	483831H1	1283	1517
4	197170.1.oct	5108547H1	1287	1423
4	197170.1.oct	g2056134	14	522
4	197170.1.oct	2182319H1	40	232
4	197170.1.oct	3187785H1	58	365
4	197170.1.oct	3538506H1	153	378
4	197170.1.oct	3538506F6	153	533
4	197170.1.oct	2521850H1	225	433
4	197170.1.oct	6317859H1	1	256
4	197170.1.oct	2402368H1	230	487
4	197170.1.oct	5688629H1	236	490
4	197170.1.oct	6176756H1	236	512
4	197170.1.oct	3785815H1	241	485
4	197170.1.oct	g2835283	246	350
4	197170.1.oct	6179452H1	265	528
4	197170.1.oct	2758648H1	282	542
4	197170.1.oct	5901704H1	331	423
4	197170.1.oct	3616975H1	331	624

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
5	345638.1.oct	3524936H1	1221	1489
5	345638.1.oct	5395774T1	1288	1796
5	345638.1.oct	2452828F6	1197	1623
5	345638.1.oct	1439670T6	1289	1695
5	345638.1.oct	2452828H1	1197	1431
5	345638.1.oct	5539601H2	1215	1439
5	345638.1.oct	2270342R6	1338	1690
5	345638.1.oct	2270342H1	1338	1591
5	345638.1.oct	2270342T6	1338	1642
5	345638.1.oct	g4175458	1372	1736
5	345638.1.oct	g1154300	1380	1548
5	345638.1.oct	g2241560	1399	1691
5	345638.1.oct	g2318255	1412	1690
5	345638.1.oct	2671159T6	1467	1891
5	345638.1.oct	g703628	1476	1696
5	345638.1.oct	5588741H1	1523	1721
5	345638.1.oct	2968103H2	1624	1926
5	345638.1.oct	2825566H1	1634	1935
5	345638.1.oct	g765459	1656	1851
5	345638.1.oct	g982043	1667	1843
5	345638.1.oct	g3872249	1790	2138
5	345638.1.oct	g3934859	1821	2138
5	345638.1.oct	2160009F6	1824	2138
5	345638.1.oct	2160081H1	1824	2069
5	345638.1.oct	2842652H1	1862	2090
5	345638.1.oct	1980041R6	1865	2138
5	345638.1.oct	1980041H1	1865	2112
5	345638.1.oct	3642137H1	1	215
5	345638.1.oct	1439670H1	1	263
5	345638.1.oct	1438620H1	1	250
5	345638.1.oct	1439670F6	1	470
5	345638.1.oct	1438620F1	1	417
5	345638.1.oct	4509739H1	49	297
5	345638.1.oct	3361128H1	68	328
5	345638.1.oct	4977040H1	69	322
5	345638.1.oct	3471990H1	69	309
5	345638.1.oct	5863955H1	71	314
5	345638.1.oct	3074876H1	72	344
5	345638.1.oct	3764710H1	82	291
5	345638.1.oct	269902H1	104	450
5	345638.1.oct	g4332349	182	516
5	345638.1.oct	1953604H1	248	464
5	345638.1.oct	5988669H1	271	460
5	345638.1.oct	4970243H1	275	547
5	345638.1.oct	3519477H1	295	462
5	345638.1.oct	4970569H1	343	602
5	345638.1.oct	4598609H1	405	587
5	345638.1.oct	2671159H1	485	729
5	345638.1.oct	2671159F6	485	927
5	345638.1.oct	2671152H1	485	729

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
5	345638.1.oct	2275853H1	505	716
5	345638.1.oct	2509419T6	508	889
5	345638.1.oct	2509419H1	515	754
5	345638.1.oct	2509419F6	515	787
5	345638.1.oct	4088123H1	529	799
5	345638.1.oct	5369561H1	918	1159
5	345638.1.oct	g2219914	953	1178
5	345638.1.oct	5607505H1	1010	1237
5	345638.1.oct	5559742H1	1021	1275
5	345638.1.oct	778081H1	1034	1244
5	345638.1.oct	3593225H1	1081	1375
5	345638.1.oct	5395774H1	579	813
5	345638.1.oct	3227545H1	663	891
5	345638.1.oct	5951114H1	817	1146
5	345638.1.oct	5951256H1	817	1145
5	345638.1.oct	2365252H1	849	928
5	345638.1.oct	4728612H1	899	1189
5	345638.1.oct	g2835252	1874	2138
5	345638.1.oct	g703518	1926	2153
5	345638.1.oct	3517178H1	1951	2073
5	345638.1.oct	g1139412	2005	2131
5	345638.1.oct	5265442H1	2006	2185
6	408784.1.dec	g4525629	1	193
6	408784.1.dec	g4113520	1	353
6	408784.1.dec	g1873896	1	272
6	408784.1.dec	g4372237	1	389
6	408784.1.dec	5426002F6	1	317
6	408784.1.dec	5426002H1	1	253
6	408784.1.dec	6264541H1	41	268
6	408784.1.dec	6566729H1	58	397
6	408784.1.dec	6569375H1	191	630
7	246526.2.dec	g3422692	1172	1278
7	246526.2.dec	g698573	1173	1273
7	246526.2.dec	2905036H1	1213	1446
7	246526.2.dec	5314181H1	1215	1363
7	246526.2.dec	g1010382	1243	1518
7	246526.2.dec	2617856H1	1249	1492
7	246526.2.dec	1574434T6	1282	1812
7	246526.2.dec	3083880H1	1315	1444
7	246526.2.dec	981687H1	1315	1544
7	246526.2.dec	1400468H1	1320	1574
7	246526.2.dec	5137157H1	1323	1589
7	246526.2.dec	g883275	1331	1633
7	246526.2.dec	g1017973	1335	1672
7	246526.2.dec	g3360494	1336	2634
7	246526.2.dec	g981374	1336	1704
7	246526.2.dec	g776347	1386	1770
7	246526.2.dec	4568542H1	1385	1573
7	246526.2.dec	g2035159	1393	1656
7	246526.2.dec	1861916F6	1419	1979

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
7	246526.2.dec	1861916H1	1419	1695
7	246526.2.dec	g574012	1420	1616
7	246526.2.dec	4591933H1	1448	1710
7	246526.2.dec	1861162T6	1467	1863
7	246526.2.dec	1861162F6	1474	1901
7	246526.2.dec	1861162H1	1475	1798
7	246526.2.dec	3856851H1	1478	1761
7	246526.2.dec	5597738H1	1495	1704
7	246526.2.dec	5919136H1	1509	1777
7	246526.2.dec	2616733H1	1510	1748
7	246526.2.dec	g3228879	1517	1913
7	246526.2.dec	1363803F1	1544	1994
7	246526.2.dec	1363803H1	1544	1791
7	246526.2.dec	g1379338	1572	1905
7	246526.2.dec	g2341495	1580	1906
7	246526.2.dec	4854205H1	1588	1848
7	246526.2.dec	358373H1	1590	1808
7	246526.2.dec	4793209H1	1600	1887
7	246526.2.dec	g1009757	1605	1748
7	246526.2.dec	4836901H1	1607	1888
7	246526.2.dec	6603577H1	1631	2157
7	246526.2.dec	5294946H1	1648	1893
7	246526.2.dec	2289413H1	1662	1880
7	246526.2.dec	2749412H1	1676	1915
7	246526.2.dec	5114945H1	1689	1960
7	246526.2.dec	4223825H1	1696	1996
7	246526.2.dec	4220586H1	1698	1962
7	246526.2.dec	1611734H1	1712	1923
7	246526.2.dec	g847365	1729	2060
7	246526.2.dec	g844344	1734	2069
7	246526.2.dec	g783315	1734	1983
7	246526.2.dec	6321704H1	1734	1933
7	246526.2.dec	4161027H1	1756	2045
7	246526.2.dec	658192H1	1760	2002
7	246526.2.dec	g2027049	1762	2050
7	246526.2.dec	1931913T6	1774	1848
7	246526.2.dec	1482438H1	1781	1980
7	246526.2.dec	1647267F6	1781	2251
7	246526.2.dec	1647343H1	1781	2022
7	246526.2.dec	5853125H1	1791	2045
7	246526.2.dec	g1231286	1793	1908
7	246526.2.dec	2425258H1	1799	2041
7	246526.2.dec	3719040H1	1806	2061
7	246526.2.dec	1494991H1	1819	2038
7	246526.2.dec	g1243109	1824	2193
7	246526.2.dec	g890161	1843	2149
7	246526.2.dec	4583873H1	1853	1995
7	246526.2.dec	2129527H1	1862	2131
7	246526.2.dec	4654582H1	1862	2124
7	246526.2.dec	g893529	1861	2146

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
7	246526.2.dec	5548011H1	1877	2174
7	246526.2.dec	2289736H1	1892	2120
7	246526.2.dec	g783093	1892	2146
7	246526.2.dec	3627179H1	1909	2072
7	246526.2.dec	g1969337	1929	2200
7	246526.2.dec	g713248	1930	2224
7	246526.2.dec	g760987	1930	2223
7	246526.2.dec	g759715	1930	2128
7	246526.2.dec	g712677	1930	2066
7	246526.2.dec	3877847H1	1931	2040
7	246526.2.dec	1972266H1	1943	2200
7	246526.2.dec	g1331147	1961	2293
7	246526.2.dec	5024160H1	1968	2243
7	246526.2.dec	1647267T6	1985	2601
7	246526.2.dec	3666919H1	2009	2108
7	246526.2.dec	3083009H1	2013	2328
7	246526.2.dec	1861916T6	2022	2595
7	246526.2.dec	2635842H1	2027	2267
7	246526.2.dec	397443T6	2036	2604
7	246526.2.dec	758011H1	2076	2382
7	246526.2.dec	838848H1	2090	2223
7	246526.2.dec	5016390H1	2099	2373
7	246526.2.dec	2822525T6	2125	2609
7	246526.2.dec	2197506F6	2125	2630
7	246526.2.dec	2197506T6	2126	2604
7	246526.2.dec	2197506H1	2125	2388
7	246526.2.dec	1722149T6	2129	2604
7	246526.2.dec	789184H1	2131	2363
7	246526.2.dec	1722149F6	2147	2578
7	246526.2.dec	1722149H1	2147	2360
7	246526.2.dec	g3278490	1	326
7	246526.2.dec	g2834735	1	67
7	246526.2.dec	g1898302	1	297
7	246526.2.dec	1495040H1	1	239
7	246526.2.dec	g4188207	9	463
7	246526.2.dec	g5435815	9	468
7	246526.2.dec	1394569H1	9	247
7	246526.2.dec	2586482H1	17	247
7	246526.2.dec	2822525F6	18	467
7	246526.2.dec	2822525H1	18	231
7	246526.2.dec	2586451H1	17	271
7	246526.2.dec	2173361H1	24	286
7	246526.2.dec	g2900274	55	484
7	246526.2.dec	g2787983	55	333
7	246526.2.dec	g2752379	73	424
7	246526.2.dec	g2816800	73	321
7	246526.2.dec	g2910688	73	176
7	246526.2.dec	3493568H1	150	414
7	246526.2.dec	1951947H1	152	275
7	246526.2.dec	1698139H1	156	352

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
7	246526.2.dec	6164569H1	156	461
7	246526.2.dec	1317624H1	157	466
7	246526.2.dec	1264328H1	173	406
7	246526.2.dec	5676507H1	181	451
7	246526.2.dec	1574434H1	193	416
7	246526.2.dec	1574566H1	193	307
7	246526.2.dec	1574582H1	193	306
7	246526.2.dec	g2883855	194	339
7	246526.2.dec	4531546H1	218	468
7	246526.2.dec	827733R1	219	698
7	246526.2.dec	2643382H1	219	447
7	246526.2.dec	827733H1	220	460
7	246526.2.dec	3397104H1	248	494
7	246526.2.dec	g2035684	249	500
7	246526.2.dec	155203H1	250	456
7	246526.2.dec	079076R6	336	782
7	246526.2.dec	079076H1	336	511
7	246526.2.dec	582057H1	376	632
7	246526.2.dec	583160H1	376	628
7	246526.2.dec	1929684F6	405	864
7	246526.2.dec	1929684H1	405	652
7	246526.2.dec	1751939H1	423	671
7	246526.2.dec	3441286H1	432	667
7	246526.2.dec	716326H1	455	596
7	246526.2.dec	2666560H1	471	719
7	246526.2.dec	g2180026	470	845
7	246526.2.dec	2479324H1	548	786
7	246526.2.dec	2479137H1	548	780
7	246526.2.dec	397443R6	559	1154
7	246526.2.dec	5519145H1	566	713
7	246526.2.dec	6615322H1	603	1130
7	246526.2.dec	g1950420	645	919
7	246526.2.dec	1929684T6	647	1223
7	246526.2.dec	3926385H1	662	936
7	246526.2.dec	4321894H1	662	921
7	246526.2.dec	079076T6	670	1236
7	246526.2.dec	g764005	752	1013
7	246526.2.dec	g703547	752	976
7	246526.2.dec	2188887F6	762	1178
7	246526.2.dec	1240601H1	762	1034
7	246526.2.dec	2059609H1	762	1015
7	246526.2.dec	2188887H1	762	1015
7	246526.2.dec	2059609R6	763	914
7	246526.2.dec	1228511H1	765	1009
7	246526.2.dec	1228592H1	765	1006
7	246526.2.dec	2173361T6	773	1232
7	246526.2.dec	1592372H1	783	907
7	246526.2.dec	2693343H1	784	1031
7	246526.2.dec	401540H1	788	926
7	246526.2.dec	g1517119	788	1107

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
7	246526.2.dec	1348393H1	801	1020
7	246526.2.dec	g4896130	813	1276
7	246526.2.dec	g3078088	814	1280
7	246526.2.dec	g3899643	818	1271
7	246526.2.dec	g5393483	824	1271
7	246526.2.dec	g704340	825	1008
7	246526.2.dec	g1014339	825	1102
7	246526.2.dec	g4690086	830	1272
7	246526.2.dec	g3899645	834	1272
7	246526.2.dec	g1955328	834	1045
7	246526.2.dec	g3245222	838	1272
7	246526.2.dec	g2238047	866	1275
7	246526.2.dec	5951508H1	869	1197
7	246526.2.dec	5949994H1	869	1125
7	246526.2.dec	5949657H1	869	1168
7	246526.2.dec	5949857H1	869	1027
7	246526.2.dec	5950094H1	869	1025
7	246526.2.dec	g2728632	872	1280
7	246526.2.dec	g2458193	890	1272
7	246526.2.dec	1931913F6	906	1284
7	246526.2.dec	1931913H1	906	1167
7	246526.2.dec	g1219072	921	1270
7	246526.2.dec	817442H1	929	1171
7	246526.2.dec	030658H1	938	1110
7	246526.2.dec	032501H1	938	1203
7	246526.2.dec	g763947	956	1260
7	246526.2.dec	g4990684	960	1275
7	246526.2.dec	g704341	970	1275
7	246526.2.dec	g1516455	974	1271
7	246526.2.dec	g2842365	976	1277
7	246526.2.dec	g5540637	983	1272
7	246526.2.dec	3617427H1	993	1305
7	246526.2.dec	g5446082	992	1274
7	246526.2.dec	g2242042	999	1267
7	246526.2.dec	g5639130	1002	1273
7	246526.2.dec	g4089555	1004	1274
7	246526.2.dec	6495914H1	1016	1471
7	246526.2.dec	271761H1	1022	1253
7	246526.2.dec	2354615H1	1039	1260
7	246526.2.dec	g3154599	1039	1434
7	246526.2.dec	6313728H1	1043	1488
7	246526.2.dec	562956H1	1046	1270
7	246526.2.dec	562956R6	1046	1268
7	246526.2.dec	500870H1	1046	1246
7	246526.2.dec	562956T6	1046	1230
7	246526.2.dec	g5638746	1055	1272
7	246526.2.dec	g1274236	1096	1536
7	246526.2.dec	3573608H1	1133	1439
7	246526.2.dec	3567436H1	1155	1309
7	246526.2.dec	1579505H1	1159	1355

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
7	246526.2.dec	1579505F6	1159	1286
7	246526.2.dec	g1331027	2169	2640
7	246526.2.dec	g4533116	2174	2642
7	246526.2.dec	g3431339	2175	2634
7	246526.2.dec	g5392578	2188	2645
7	246526.2.dec	2188887T6	2189	2595
7	246526.2.dec	3702561H1	2206	2518
7	246526.2.dec	3432072H1	2205	2377
7	246526.2.dec	g4076803	2211	2643
7	246526.2.dec	g5395302	2213	2649
7	246526.2.dec	g4003630	2217	2643
7	246526.2.dec	g5631457	2220	2640
7	246526.2.dec	g3870470	2222	2643
7	246526.2.dec	g3899705	2229	2642
7	246526.2.dec	2021042H1	2232	2438
7	246526.2.dec	6405258H2	2235	2421
7	246526.2.dec	6156086H1	2240	2547
7	246526.2.dec	g713604	2247	2642
7	246526.2.dec	g4282562	2244	2634
7	246526.2.dec	g2752080	2244	2406
7	246526.2.dec	g1331098	2251	2659
7	246526.2.dec	1496240H1	2252	2473
7	246526.2.dec	g1224147	2253	2643
7	246526.2.dec	1496240T1	2252	2602
7	246526.2.dec	g3801920	2262	2648
7	246526.2.dec	2429495H1	2265	2499
7	246526.2.dec	g3094872	2269	2643
7	246526.2.dec	g1010383	2272	2640
7	246526.2.dec	g3701189	2279	2648
7	246526.2.dec	g2674626	2284	2646
7	246526.2.dec	g981375	2286	2642
7	246526.2.dec	g2279829	2294	2633
7	246526.2.dec	g883276	2302	2654
7	246526.2.dec	g3742404	2311	2644
7	246526.2.dec	6361538H2	2313	2431
7	246526.2.dec	g1219974	2315	2642
7	246526.2.dec	g1201439	2319	2648
7	246526.2.dec	g723228	2325	2645
7	246526.2.dec	g898941	2326	2629
7	246526.2.dec	g760988	2354	2635
7	246526.2.dec	g2659181	2353	2645
7	246526.2.dec	g2020840	2355	2643
7	246526.2.dec	g2559564	2358	2649
7	246526.2.dec	g846893	2365	2643
7	246526.2.dec	1986742H1	2364	2558
7	246526.2.dec	g2969330	2368	2639
7	246526.2.dec	g1018535	2384	2608
7	246526.2.dec	g782865	2387	2642
7	246526.2.dec	g566344	2388	2612
7	246526.2.dec	1598359H1	2413	2622

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
7	246526.2.dec	1598358H1	2413	2623
7	246526.2.dec	1679021H1	2416	2628
7	246526.2.dec	2696192H1	2425	2634
7	246526.2.dec	g1017889	2427	2644
7	246526.2.dec	2364484H1	2433	2646
7	246526.2.dec	g3092039	2445	2645
7	246526.2.dec	g704228	2444	2620
7	246526.2.dec	g2213054	2495	2643
7	246526.2.dec	g3840446	2494	2642
7	246526.2.dec	2770933H1	2517	2634
7	246526.2.dec	4830839H1	2526	2650
7	246526.2.dec	289634H1	2534	2634
7	246526.2.dec	1358265H1	2537	2815
7	246526.2.dec	g2837523	2592	2965
8	200488.5.dec	4043361H1	1	265
8	200488.5.dec	4043361F6	1	571
8	200488.5.dec	5400109H1	38	168
8	200488.5.dec	5620913H1	45	321
8	200488.5.dec	572762H1	45	303
8	200488.5.dec	g680776	46	174
8	200488.5.dec	6437310H1	55	635
8	200488.5.dec	4043361T6	118	718
8	200488.5.dec	g1618321	320	699
8	200488.5.dec	4880281H1	523	754
8	200488.5.dec	5949678H1	525	771
9	474878.1.dec	571127H1	1497	1715
9	474878.1.dec	2328233H1	1525	1780
9	474878.1.dec	2328233R6	1525	2032
9	474878.1.dec	g1940321	1531	1844
9	474878.1.dec	6157851H1	1534	1697
9	474878.1.dec	g3742402	1539	1852
9	474878.1.dec	g1940948	1544	1721
9	474878.1.dec	g3166966	1551	1979
9	474878.1.dec	6157772H1	1567	1806
9	474878.1.dec	778655H1	1578	1821
9	474878.1.dec	1581071H1	1581	1781
9	474878.1.dec	5099087H1	1590	1857
9	474878.1.dec	1314472H1	1591	1864
9	474878.1.dec	1784735H1	1595	1843
9	474878.1.dec	584498H1	1596	1932
9	474878.1.dec	1344492H1	1609	1851
9	474878.1.dec	4068471H1	1634	1802
9	474878.1.dec	506313H1	1678	1902
9	474878.1.dec	6108290H1	1679	1952
9	474878.1.dec	2097212H1	1685	1874
9	474878.1.dec	6485186H1	1688	2236
9	474878.1.dec	6266715H1	1690	2262
9	474878.1.dec	745544R6	1690	2036
9	474878.1.dec	745544H1	1691	1955
9	474878.1.dec	4882217H1	1715	1999

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
9	474878.1.dec	2014391H1	1723	1979
9	474878.1.dec	1979002H1	1754	2044
9	474878.1.dec	3780778H1	1765	2080
9	474878.1.dec	1383473T6	1767	2386
9	474878.1.dec	1912211H1	1784	2036
9	474878.1.dec	4953404H1	1787	2041
9	474878.1.dec	3593485H1	1786	2082
9	474878.1.dec	1704049H1	1799	2017
9	474878.1.dec	532564H1	1800	2020
9	474878.1.dec	2460229H1	1808	2034
9	474878.1.dec	3122555H1	1810	2116
9	474878.1.dec	1007753H1	1826	2127
9	474878.1.dec	4550034T1	1830	2386
9	474878.1.dec	745544T6	1843	2382
9	474878.1.dec	g749175	1851	2125
9	474878.1.dec	2201650H1	1852	2109
9	474878.1.dec	6357768H1	1860	1983
9	474878.1.dec	1906164T6	1895	2396
9	474878.1.dec	2082940H1	1899	2137
9	474878.1.dec	2081388H1	1899	2137
9	474878.1.dec	g1383678	1912	2339
9	474878.1.dec	1298690H1	1913	2163
9	474878.1.dec	1298690F1	1914	2323
9	474878.1.dec	6326213H1	1916	2217
9	474878.1.dec	4979876H1	1937	2206
9	474878.1.dec	839539H1	1938	2148
9	474878.1.dec	g2913007	1943	2424
9	474878.1.dec	6430283H1	1948	2410
9	474878.1.dec	g3003791	1946	2425
9	474878.1.dec	2302122T6	1962	2388
9	474878.1.dec	1855791T6	1964	2379
9	474878.1.dec	3178987H1	1970	2280
9	474878.1.dec	6321190H1	1972	2245
9	474878.1.dec	3482645T6	1984	2387
9	474878.1.dec	2260661T6	1991	2390
9	474878.1.dec	6326393H1	1995	2282
9	474878.1.dec	g3418836	2004	2425
9	474878.1.dec	g5659013	2008	2425
9	474878.1.dec	g3900513	2009	2425
9	474878.1.dec	2328233T6	2008	2387
9	474878.1.dec	g3988728	2012	2427
9	474878.1.dec	g3095324	2013	2432
9	474878.1.dec	g4072902	2016	2425
9	474878.1.dec	g1727223	2023	2425
9	474878.1.dec	g3597743	2024	2426
9	474878.1.dec	g2554182	2029	2426
9	474878.1.dec	g2397860	2031	2426
9	474878.1.dec	g1940832	2031	2425
9	474878.1.dec	334239H1	2033	2262
9	474878.1.dec	g1941153	2037	2425

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
9	474878.1.dec	g2465956	2046	2426
9	474878.1.dec	g3002018	2055	2426
9	474878.1.dec	g3052386	2060	2429
9	474878.1.dec	g658645	2060	2425
9	474878.1.dec	5021683H1	2061	2342
9	474878.1.dec	4646243H1	2062	2326
9	474878.1.dec	5021683T1	2064	2379
9	474878.1.dec	073688H1	2071	2326
9	474878.1.dec	073962H1	2071	2213
9	474878.1.dec	g1689373	2076	2413
9	474878.1.dec	g988331	2101	2435
9	474878.1.dec	g2183351	2100	2425
9	474878.1.dec	271231H1	2100	2298
9	474878.1.dec	g989410	2108	2419
9	474878.1.dec	g2047031	2107	2425
9	474878.1.dec	4219112H1	2108	2374
9	474878.1.dec	g564678	2112	2425
9	474878.1.dec	4202793H1	2125	2410
9	474878.1.dec	g2835238	2125	2425
9	474878.1.dec	1576547H1	2125	2289
9	474878.1.dec	g1312586	2129	2426
9	474878.1.dec	g4986384	2154	2429
9	474878.1.dec	780421H1	2162	2292
9	474878.1.dec	g749280	2167	2427
9	474878.1.dec	4717682H1	1	248
9	474878.1.dec	2604722H1	18	247
9	474878.1.dec	2260661R6	22	421
9	474878.1.dec	2260661H1	22	274
9	474878.1.dec	4795373H1	22	272
9	474878.1.dec	3102252H1	27	352
9	474878.1.dec	5037329H1	30	295
9	474878.1.dec	4530781H1	30	285
9	474878.1.dec	5400139H1	32	178
9	474878.1.dec	6604055H1	44	451
9	474878.1.dec	1258366H1	34	246
9	474878.1.dec	3510895H1	42	314
9	474878.1.dec	3465753H1	45	379
9	474878.1.dec	4114074H1	52	167
9	474878.1.dec	3649065H1	62	166
9	474878.1.dec	6138239H1	68	366
9	474878.1.dec	3750720H1	75	197
9	474878.1.dec	2960324H1	84	242
9	474878.1.dec	g651892	187	490
9	474878.1.dec	4989817H1	212	495
9	474878.1.dec	3482645F6	303	867
9	474878.1.dec	3482645H1	303	498
9	474878.1.dec	1395412H1	336	595
9	474878.1.dec	5393935H1	385	650
9	474878.1.dec	4797624H1	396	665
9	474878.1.dec	6541193H1	414	931

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
9	474878.1.dec	2490605H1	449	689
9	474878.1.dec	1344946F6	476	921
9	474878.1.dec	g5233250	495	696
9	474878.1.dec	3504674H1	497	772
9	474878.1.dec	3748060H1	512	807
9	474878.1.dec	g1727282	550	722
9	474878.1.dec	352603H1	629	831
9	474878.1.dec	2302122R6	636	1097
9	474878.1.dec	2302122H1	636	870
9	474878.1.dec	4177761H1	647	919
9	474878.1.dec	190033H1	673	890
9	474878.1.dec	359191H1	673	896
9	474878.1.dec	1855791F6	675	1208
9	474878.1.dec	1855791H1	675	874
9	474878.1.dec	4649538H1	723	971
9	474878.1.dec	5404524H1	732	873
9	474878.1.dec	g658646	828	1137
9	474878.1.dec	1906164F6	860	1103
9	474878.1.dec	1906164H1	860	961
9	474878.1.dec	5120370H1	901	1016
9	474878.1.dec	5120063H1	901	1198
9	474878.1.dec	022665H1	957	1302
9	474878.1.dec	2185728H1	975	1241
9	474878.1.dec	4886861H1	976	1278
9	474878.1.dec	6159129H1	982	1209
9	474878.1.dec	5104424H1	994	1263
9	474878.1.dec	010115H1	1008	1346
9	474878.1.dec	5048787H1	1018	1268
9	474878.1.dec	593263H1	1018	1249
9	474878.1.dec	g824550	1019	1318
9	474878.1.dec	5188638H1	1023	1338
9	474878.1.dec	1383473F6	1069	1564
9	474878.1.dec	1383473H1	1069	1312
9	474878.1.dec	1381362H1	1069	1297
9	474878.1.dec	3518142H1	1079	1403
9	474878.1.dec	5661759H1	1081	1343
9	474878.1.dec	4550002H1	1117	1352
9	474878.1.dec	713452H1	1133	1321
9	474878.1.dec	g668336	1149	1410
9	474878.1.dec	g573132	1149	1472
9	474878.1.dec	g696465	1150	1533
9	474878.1.dec	1898273H1	1196	1457
9	474878.1.dec	g988498	1197	1520
9	474878.1.dec	3525794H1	1198	1470
9	474878.1.dec	2266314H1	1199	1443
9	474878.1.dec	5175326H1	1211	1421
9	474878.1.dec	5597046H1	1270	1537
9	474878.1.dec	6514953H1	1272	1777
9	474878.1.dec	g1941526	1281	1686
9	474878.1.dec	1924150H1	1305	1544

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
9	474878.1.dec	4900790H1	1340	1520
9	474878.1.dec	1816879H1	1347	1608
9	474878.1.dec	5836742H1	1370	1643
9	474878.1.dec	2265560H1	1375	1637
9	474878.1.dec	3702961H1	1380	1675
9	474878.1.dec	1907523H1	1440	1684
9	474878.1.dec	g5100926	1444	1854
9	474878.1.dec	6551418H1	1446	2002
9	474878.1.dec	3958231H2	1463	1738
9	474878.1.dec	4543030H1	1484	1564
9	474878.1.dec	2801727H1	1493	1695
9	474878.1.dec	g3174029	2178	2429
9	474878.1.dec	g644901	2178	2425
9	474878.1.dec	408264H1	2186	2420
9	474878.1.dec	1417510H1	2187	2425
9	474878.1.dec	g4194592	2192	2429
9	474878.1.dec	1547442H1	2197	2378
9	474878.1.dec	g1383626	2210	2430
9	474878.1.dec	6569854H1	2225	2427
9	474878.1.dec	g824551	2233	2435
9	474878.1.dec	2366889H1	2272	2425
9	474878.1.dec	2371405H1	2272	2425
9	474878.1.dec	3481366H1	2278	2431
9	474878.1.dec	g4438805	2293	2425
10	335916.2.dec	6497614H1	1	488
10	335916.2.dec	6457162H1	45	466
10	335916.2.dec	3110489H1	81	346
10	335916.2.dec	2782031F6	241	655
10	335916.2.dec	2782031H1	241	498
10	335916.2.dec	g4622020	348	594
10	335916.2.dec	2508394F6	377	743
10	335916.2.dec	2508394H1	377	623
10	335916.2.dec	3253880H1	393	636
10	335916.2.dec	4753078H1	450	560
10	335916.2.dec	2664350H1	617	830
10	335916.2.dec	5841727H1	708	965
10	335916.2.dec	3345808H1	758	853
10	335916.2.dec	1664667F6	790	1241
10	335916.2.dec	1664667H1	790	1036
10	335916.2.dec	6495655H1	905	1341
10	335916.2.dec	1730823H1	960	1040
10	335916.2.dec	3294429H1	969	1226
10	335916.2.dec	3371239H1	1019	1259
10	335916.2.dec	2861953H1	1053	1330
10	335916.2.dec	3401194H1	1053	1278
10	335916.2.dec	2861953F6	1053	1558
10	335916.2.dec	3257620H1	1113	1251
10	335916.2.dec	867163H1	1133	1372
10	335916.2.dec	867163R6	1133	1401
10	335916.2.dec	g2324543	1269	1620

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
10	335916.2.dec	1627014H1	1488	1720
10	335916.2.dec	g1939354	1506	1766
10	335916.2.dec	g2107851	1506	1858
10	335916.2.dec	912981H1	1615	1747
10	335916.2.dec	2111286H1	1616	1863
10	335916.2.dec	3790008H1	1693	1809
10	335916.2.dec	1214420H1	1700	1936
10	335916.2.dec	3535232H1	1798	2072
10	335916.2.dec	3257037H1	1810	2064
10	335916.2.dec	3210776H1	1820	2024
11	040422.12.dec	3343947H1	1	210
11	040422.12.dec	3343947F6	1	398
11	040422.12.dec	4183830H1	23	207
11	040422.12.dec	4792750H1	25	295
11	040422.12.dec	3159520H1	27	304
11	040422.12.dec	3296383H1	28	279
11	040422.12.dec	5197324H1	29	284
11	040422.12.dec	5197324F6	29	299
11	040422.12.dec	g3341989	42	1400
11	040422.12.dec	5978581H1	51	292
11	040422.12.dec	3898429H1	52	272
11	040422.12.dec	5605234H1	53	276
11	040422.12.dec	5302780H1	53	291
11	040422.12.dec	3592605H1	64	359
11	040422.12.dec	3593031H1	64	368
11	040422.12.dec	g1727841	70	483
11	040422.12.dec	6552493H1	112	701
11	040422.12.dec	6557908H1	112	591
11	040422.12.dec	4051117H1	224	509
11	040422.12.dec	3293317H1	477	729
11	040422.12.dec	2928983H1	499	798
11	040422.12.dec	3032868H1	515	808
11	040422.12.dec	5570320H1	666	831
11	040422.12.dec	4418383H1	691	897
11	040422.12.dec	4747915H1	718	987
11	040422.12.dec	g3096317	720	1177
11	040422.12.dec	3343947T6	725	1356
11	040422.12.dec	4371455H1	742	1023
11	040422.12.dec	1978317T6	801	1359
11	040422.12.dec	1978317R6	811	1196
11	040422.12.dec	1978317H1	811	1110
11	040422.12.dec	g3037965	922	1400
11	040422.12.dec	482585T6	937	1380
11	040422.12.dec	5029812H1	937	1150
11	040422.12.dec	658005H1	937	1115
11	040422.12.dec	1610157T1	937	995
11	040422.12.dec	1610157T6	941	1360
11	040422.12.dec	g1046767	974	1300
11	040422.12.dec	g5113655	983	1401
11	040422.12.dec	g2161987	985	1403

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
11	040422.12.dec	g1046664	1016	1400
11	040422.12.dec	4146468H1	1040	1288
11	040422.12.dec	g1727662	1053	1397
11	040422.12.dec	g4149302	1084	1401
11	040422.12.dec	3219363H1	1117	1394
11	040422.12.dec	g3871333	1123	1399
11	040422.12.dec	g2784520	1173	1409
11	040422.12.dec	2346542F6	1191	1400
11	040422.12.dec	2346542H1	1191	1421
11	040422.12.dec	g3037903	1314	1400
12	977651.2.dec	2801809H1	207	469
12	977651.2.dec	g1267440	205	617
12	977651.2.dec	2910841H1	209	468
12	977651.2.dec	4045484H1	211	488
12	977651.2.dec	4639491H1	213	471
12	977651.2.dec	2182080H1	231	511
12	977651.2.dec	3154813H1	246	504
12	977651.2.dec	3873262H1	300	562
12	977651.2.dec	1459945H1	300	540
12	977651.2.dec	4635570H1	309	553
12	977651.2.dec	3254753H1	315	573
12	977651.2.dec	986038H1	339	571
12	977651.2.dec	1541940H1	350	575
12	977651.2.dec	4466419H1	356	598
12	977651.2.dec	4466417H1	359	606
12	977651.2.dec	151211H1	372	604
12	977651.2.dec	4981429H1	381	635
12	977651.2.dec	g728181	395	643
12	977651.2.dec	1748157H1	397	680
12	977651.2.dec	4219930H1	402	701
12	977651.2.dec	1574381H1	402	633
12	977651.2.dec	1574381F6	402	660
12	977651.2.dec	4464523H1	412	598
12	977651.2.dec	4906687H2	433	674
12	977651.2.dec	4044623H1	438	705
12	977651.2.dec	3037055H1	451	734
12	977651.2.dec	1977380H1	452	679
12	977651.2.dec	2115723H1	455	567
12	977651.2.dec	5185327H1	459	693
12	977651.2.dec	907607H1	529	680
12	977651.2.dec	g3076955	621	1092
12	977651.2.dec	g2946373	634	1091
12	977651.2.dec	g4003859	654	1096
12	977651.2.dec	g4893606	682	1091
12	977651.2.dec	g3400779	696	1099
12	977651.2.dec	g3214725	697	1091
12	977651.2.dec	g2537968	705	1096
12	977651.2.dec	g5152617	720	1092
12	977651.2.dec	g3597940	727	1105
12	977651.2.dec	g2670173	775	1090

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
12	977651.2.dec	g4072011	812	1094
12	977651.2.dec	g1802488	816	1093
12	977651.2.dec	g4195857	822	1096
12	977651.2.dec	g4195469	898	1094
12	977651.2.dec	1921146R6	1	443
12	977651.2.dec	1921146H1	1	216
12	977651.2.dec	827751H1	1	274
12	977651.2.dec	2294102H1	75	353
12	977651.2.dec	2786002H1	77	349
12	977651.2.dec	6357235H1	160	372
12	977651.2.dec	2495317H1	161	483
12	977651.2.dec	3648560H1	169	343
12	977651.2.dec	855978H1	169	401
12	977651.2.dec	2409671H1	169	410
12	977651.2.dec	5021802H1	169	446
12	977651.2.dec	4166987H1	170	286
12	977651.2.dec	2814612H1	169	487
12	977651.2.dec	2578951H1	174	383
12	977651.2.dec	6171520H1	174	467
12	977651.2.dec	3347246H1	175	433
12	977651.2.dec	3491814H1	178	448
12	977651.2.dec	3360455H1	182	466
12	977651.2.dec	5863129H1	189	248
12	977651.2.dec	4725591H1	192	453
12	977651.2.dec	4798920H1	192	425
12	977651.2.dec	4725558H1	192	407
12	977651.2.dec	g1012357	194	481
12	977651.2.dec	g4680704	194	1096
12	977651.2.dec	2793693F6	199	606
12	977651.2.dec	2793693H1	199	491
12	977651.2.dec	2860948H1	199	459
12	977651.2.dec	2760646H1	199	450
12	977651.2.dec	2889910H1	199	382
12	977651.2.dec	4541324H1	199	465
12	977651.2.dec	3325062H1	201	458
12	977651.2.dec	4675275H1	202	303
12	977651.2.dec	3741914H1	202	502
12	977651.2.dec	2738790H1	202	440
12	977651.2.dec	4547040H1	202	357
12	977651.2.dec	4800415H1	204	492
12	977651.2.dec	3150944H1	202	377
12	977651.2.dec	g1802603	203	599
12	977651.2.dec	3050095H1	203	489
12	977651.2.dec	2635706H1	204	471
12	977651.2.dec	2452359H1	204	454
12	977651.2.dec	2692134H1	204	453
12	977651.2.dec	2545415H1	204	453
12	977651.2.dec	3523020H1	205	542
12	977651.2.dec	1919594H1	204	361
12	977651.2.dec	2780137H1	204	449

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
12	977651.2.dec	6304468H2	205	725
12	977651.2.dec	5017869H1	205	494
12	977651.2.dec	3213669H1	205	443
12	977651.2.dec	3521891H1	205	380
12	977651.2.dec	3129806H1	204	515
12	977651.2.dec	2106050H1	206	453
12	977651.2.dec	g1164443	206	469
12	977651.2.dec	3603114H1	207	514
12	977651.2.dec	2909904H1	207	479
12	977651.2.dec	4387078H1	207	467
13	012432.5.dec	2610935H1	1	244
13	012432.5.dec	712941H1	20	161
13	012432.5.dec	4175484H1	20	318
13	012432.5.dec	3458411H1	25	281
13	012432.5.dec	3286928H1	25	273
13	012432.5.dec	3297142H1	24	263
13	012432.5.dec	2665744H1	23	257
13	012432.5.dec	804988H1	25	251
13	012432.5.dec	3983449H1	22	207
13	012432.5.dec	3286928F6	25	586
13	012432.5.dec	3458411F6	25	423
13	012432.5.dec	5070204H1	26	334
13	012432.5.dec	660788H1	25	277
13	012432.5.dec	3464584H1	26	214
13	012432.5.dec	5992614H1	26	322
13	012432.5.dec	5472074H1	27	277
13	012432.5.dec	4913558H1	28	302
13	012432.5.dec	593561H1	29	183
13	012432.5.dec	2718265H1	31	276
13	012432.5.dec	3463817F6	31	519
13	012432.5.dec	3463817H1	31	327
13	012432.5.dec	194287H1	32	223
13	012432.5.dec	3391154H1	34	281
13	012432.5.dec	3391454H1	34	278
13	012432.5.dec	3375131H1	37	270
13	012432.5.dec	2920915H1	40	310
13	012432.5.dec	5163335H1	58	292
13	012432.5.dec	g3401307	57	411
13	012432.5.dec	g1807207	103	238
13	012432.5.dec	597038H1	119	311
13	012432.5.dec	4343602H1	188	470
13	012432.5.dec	1216935H1	439	590
14	059263.6.dec	g4333810	446	906
14	059263.6.dec	5907201H1	1	308
14	059263.6.dec	g1809245	88	2109
14	059263.6.dec	4178992H2	116	368
14	059263.6.dec	1467979H1	620	818
14	059263.6.dec	1467979F6	620	951
14	059263.6.dec	3085763H1	625	929
14	059263.6.dec	4326394H1	480	678

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
14	059263.6.dec	6546375H1	449	937
14	059263.6.dec	5913060H1	536	833
14	059263.6.dec	6515415H1	575	1105
14	059263.6.dec	5924357H1	616	889
14	059263.6.dec	449317T6	1613	2067
14	059263.6.dec	3845518H1	1635	1913
14	059263.6.dec	g5431445	1665	2115
14	059263.6.dec	735396H1	1304	1545
14	059263.6.dec	4709292H1	1260	1515
14	059263.6.dec	735396R1	1304	1840
14	059263.6.dec	527201H1	1336	1585
14	059263.6.dec	2755840H1	1410	1664
14	059263.6.dec	5602808H1	1412	1676
14	059263.6.dec	3449574H1	1412	1526
14	059263.6.dec	6269421H1	1419	1777
14	059263.6.dec	445961F1	1498	2109
14	059263.6.dec	3166822H1	1073	1359
14	059263.6.dec	g2013303	1007	1262
14	059263.6.dec	6269333H1	1183	1811
14	059263.6.dec	338737H1	1246	1484
14	059263.6.dec	3885430H2	1252	1506
14	059263.6.dec	3637017H1	969	1261
14	059263.6.dec	4959483H1	1006	1259
14	059263.6.dec	6437085H1	639	1152
14	059263.6.dec	3566975H1	711	955
14	059263.6.dec	445961R6	715	1248
14	059263.6.dec	3162586H1	1520	1795
14	059263.6.dec	338360H1	1533	1650
14	059263.6.dec	4367573H1	1559	1832
14	059263.6.dec	449317H1	944	1113
14	059263.6.dec	5907575H1	947	1239
14	059263.6.dec	342416H1	960	1197
14	059263.6.dec	3162459H1	947	1231
14	059263.6.dec	g560331	1887	2109
14	059263.6.dec	1519675T6	1909	2059
14	059263.6.dec	g668542	634	903
14	059263.6.dec	6430603H1	639	1124
14	059263.6.dec	g668543	634	893
14	059263.6.dec	g900542	634	946
14	059263.6.dec	445961R1	715	1205
14	059263.6.dec	512782H1	715	967
14	059263.6.dec	2431467H1	715	893
14	059263.6.dec	3242035H1	749	988
14	059263.6.dec	5913544H1	785	1063
14	059263.6.dec	4193622H1	818	1094
14	059263.6.dec	4958901H1	857	1117
14	059263.6.dec	4439155H1	871	1144
14	059263.6.dec	449788H1	944	1104
14	059263.6.dec	g775766	273	616
14	059263.6.dec	4708214H1	275	550

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
14	059263.6.dec	2151592H1	281	553
14	059263.6.dec	g814279	319	720
14	059263.6.dec	5373649H1	333	554
14	059263.6.dec	4708510H1	183	301
14	059263.6.dec	5925904H1	242	453
14	059263.6.dec	g1173538	242	1318
14	059263.6.dec	g389367	252	666
14	059263.6.dec	g1950140	255	667
14	059263.6.dec	g615808	1853	2109
14	059263.6.dec	3937956H1	1859	2071
14	059263.6.dec	2906383H1	1823	2109
14	059263.6.dec	3421560H1	1827	2083
14	059263.6.dec	g3076896	1841	2109
14	059263.6.dec	5434173H1	1768	1998
14	059263.6.dec	g2324590	1793	2109
14	059263.6.dec	g317469	1795	2109
14	059263.6.dec	3843223H1	1805	2084
14	059263.6.dec	g2269635	1756	2110
14	059263.6.dec	g2388765	1761	2109
14	059263.6.dec	g2214360	1767	2109
14	059263.6.dec	445961T6	1716	2068
14	059263.6.dec	5079421H1	1738	1847
14	059263.6.dec	4944041H1	1744	2021
14	059263.6.dec	1453911F6	1748	2019
14	059263.6.dec	3846369H1	1673	1909
14	059263.6.dec	g828896	1675	2109
14	059263.6.dec	g3870013	1678	2109
14	059263.6.dec	3002426T6	1683	2069
14	059263.6.dec	g3307105	1688	2111
14	059263.6.dec	g389366	1701	2109
14	059263.6.dec	g3834908	1690	2108
15	196774.3.dec	4198864H1	366	645
15	196774.3.dec	6543639H1	1	536
15	196774.3.dec	5467282H1	349	610
15	196774.3.dec	5467289H1	349	605
15	196774.3.dec	6545364H1	383	952
15	196774.3.dec	3124504H1	596	882
15	196774.3.dec	2858708T6	720	1100
15	196774.3.dec	1656694T6	752	1086
16	233624.11.dec	2578538F6	1	480
16	233624.11.dec	2578538H1	1	187
16	233624.11.dec	4624394H1	26	149
16	233624.11.dec	2478423H1	54	282
16	233624.11.dec	g1999348	59	188
16	233624.11.dec	2136789F6	288	634
16	233624.11.dec	2136789H1	288	511
16	233624.11.dec	5350727H1	399	563
16	233624.11.dec	5350889H1	399	524
16	233624.11.dec	3639605H1	581	871
16	233624.11.dec	3765020H1	612	906

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
17	228585.3.dec	1740045R6	1839	2332
17	228585.3.dec	1739439H1	1839	2088
17	228585.3.dec	1740045H1	1839	2053
17	228585.3.dec	5849788H1	1839	1981
17	228585.3.dec	5374048H1	1849	2100
17	228585.3.dec	4723812H1	1855	2126
17	228585.3.dec	2288963H1	1877	2123
17	228585.3.dec	1373365H1	1895	2124
17	228585.3.dec	1595644F6	1900	2332
17	228585.3.dec	4341071H1	1900	2206
17	228585.3.dec	1595644H1	1900	2114
17	228585.3.dec	2245012H1	1903	2111
17	228585.3.dec	532797T6	1927	2525
17	228585.3.dec	1400814H1	1953	2236
17	228585.3.dec	3945768H1	1974	2247
17	228585.3.dec	6307190H1	1986	2542
17	228585.3.dec	4313085H1	1998	2282
17	228585.3.dec	1595644T6	2005	2530
17	228585.3.dec	1712978T6	2037	2530
17	228585.3.dec	1412604T6	2065	2548
17	228585.3.dec	620296T6	2077	2525
17	228585.3.dec	1942348R6	2079	2588
17	228585.3.dec	4312619H1	2078	2365
17	228585.3.dec	2123967T6	2092	2532
17	228585.3.dec	663135T6	2103	2524
17	228585.3.dec	g5689560	2116	5914
17	228585.3.dec	g4685449	2116	2557
17	228585.3.dec	g4984720	2116	2535
17	228585.3.dec	2570503H1	3742	3978
17	228585.3.dec	4722814H1	3806	3916
17	228585.3.dec	1949846H1	2251	2496
17	228585.3.dec	1949815H1	2251	2496
17	228585.3.dec	5904662H1	2255	2551
17	228585.3.dec	g819594	2266	2622
17	228585.3.dec	4768762H1	2279	2556
17	228585.3.dec	g517574	2128	2616
17	228585.3.dec	6360619H1	2122	2294
17	228585.3.dec	2400848H1	2127	2343
17	228585.3.dec	620296H1	2127	2334
17	228585.3.dec	4311031H1	2127	2317
17	228585.3.dec	5902549H1	2142	2436
17	228585.3.dec	1942348H1	2127	2347
17	228585.3.dec	5659857H1	2131	2307
17	228585.3.dec	5614086H1	2142	2422
17	228585.3.dec	5898857H1	2142	2416
17	228585.3.dec	5898671H1	2142	2410
17	228585.3.dec	1673835T6	2149	2524
17	228585.3.dec	5139434H1	2145	2412
17	228585.3.dec	6131287H1	2180	2445
17	228585.3.dec	5679165H1	3595	3673

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
17	228585.3.dec	2658040F6	3621	4130
17	228585.3.dec	3762503H1	3733	3990
17	228585.3.dec	6118313H1	1	579
17	228585.3.dec	4787759H1	1255	1490
17	228585.3.dec	g990857	1308	1644
17	228585.3.dec	g946099	1308	1455
17	228585.3.dec	g878889	1308	1405
17	228585.3.dec	532797R6	1322	1701
17	228585.3.dec	532797H1	1323	1525
17	228585.3.dec	5833867H1	1348	1619
17	228585.3.dec	3726741H1	1377	1671
17	228585.3.dec	1712978F6	1476	1889
17	228585.3.dec	1712978H1	1476	1694
17	228585.3.dec	3770873H1	1489	1788
17	228585.3.dec	5876837H1	1498	1783
17	228585.3.dec	663135R6	1574	2127
17	228585.3.dec	3147094H1	165	447
17	228585.3.dec	3593276H1	210	524
17	228585.3.dec	1412604F6	352	898
17	228585.3.dec	6121951H1	1695	2233
17	228585.3.dec	5690251H1	1700	1975
17	228585.3.dec	6173835H1	1711	1942
17	228585.3.dec	g990056	1708	2017
17	228585.3.dec	3600703H1	1721	2011
17	228585.3.dec	5689188H1	1780	2048
17	228585.3.dec	2907705H1	1798	2050
17	228585.3.dec	1412604H1	352	622
17	228585.3.dec	g677056	424	633
17	228585.3.dec	g672789	430	758
17	228585.3.dec	g892790	431	666
17	228585.3.dec	g775645	431	678
17	228585.3.dec	1297889H1	541	784
17	228585.3.dec	1297889F1	541	765
17	228585.3.dec	g4069788	2121	2560
17	228585.3.dec	g564864	2122	2474
17	228585.3.dec	g671393	2122	2359
17	228585.3.dec	g518101	2116	2524
17	228585.3.dec	g3693534	2116	2454
17	228585.3.dec	g519265	2116	2375
17	228585.3.dec	g615632	2116	2306
17	228585.3.dec	g4888013	2116	2516
17	228585.3.dec	5371718H1	602	850
17	228585.3.dec	663135H1	1574	1838
17	228585.3.dec	4716435H1	1580	1675
17	228585.3.dec	5579252H1	1619	1878
17	228585.3.dec	6121671H1	1652	2080
17	228585.3.dec	2006293H1	1673	1796
17	228585.3.dec	6122051H1	1693	2025
17	228585.3.dec	3614928H1	3171	3452
17	228585.3.dec	6308868H1	3322	3848

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
17	228585.3.dec	2775486H1	3344	3559
17	228585.3.dec	6131367H1	3561	3828
17	228585.3.dec	1954291H1	2639	2859
17	228585.3.dec	5859988H1	2644	2885
17	228585.3.dec	g2050587	2626	3105
17	228585.3.dec	g314160	2454	2782
17	228585.3.dec	g891612	2514	2615
17	228585.3.dec	2291482H1	2485	2601
17	228585.3.dec	g2140727	2547	2601
17	228585.3.dec	4013255H1	2601	2886
17	228585.3.dec	g274421	4639	4951
17	228585.3.dec	1450839H1	5713	5923
17	228585.3.dec	g389990	4628	4941
17	228585.3.dec	2658040H1	3900	4130
17	228585.3.dec	g2559863	2361	2497
17	228585.3.dec	2658040T6	2416	2533
17	228585.3.dec	3313954H1	2371	2595
17	228585.3.dec	2123967H1	2335	2605
17	228585.3.dec	5911041H1	2382	2618
17	228585.3.dec	861748H1	2361	2568
17	228585.3.dec	g796237	2293	2613
17	228585.3.dec	2123967F6	2315	2607
17	228585.3.dec	g876300	2295	2615
17	228585.3.dec	4577236H1	2327	2593
17	228585.3.dec	2413167H1	1193	1436
17	228585.3.dec	6060583H1	1142	1192
17	228585.3.dec	4759715H1	1157	1421
17	228585.3.dec	6296788H1	1171	1435
17	228585.3.dec	4228967H1	1205	1467
17	228585.3.dec	4060180H1	1187	1471
17	228585.3.dec	1902883H1	905	1155
17	228585.3.dec	3761073H1	924	1219
17	228585.3.dec	g698612	1012	1235
17	228585.3.dec	5576750H1	1048	1301
17	228585.3.dec	3024344H1	1055	1318
17	228585.3.dec	g876656	1062	1415
17	228585.3.dec	4758756H1	1066	1240
17	228585.3.dec	2918456H1	704	989
17	228585.3.dec	6478720H1	735	1254
17	228585.3.dec	4136187H1	746	1024
17	228585.3.dec	g878232	782	1137
17	228585.3.dec	1673835H1	812	1048
17	228585.3.dec	1673806H1	812	1037
17	228585.3.dec	4115301H1	878	1134
18	198840.3.dec	908528H1	903	1052
18	198840.3.dec	g1228717	895	1052
18	198840.3.dec	g2719009	928	1052
18	198840.3.dec	571865H1	792	999
18	198840.3.dec	2289267H1	772	980
18	198840.3.dec	g3050309	786	974

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
18	198840.3.dec	g1188260	548	769
18	198840.3.dec	534108F1	555	1052
18	198840.3.dec	5020606T1	563	1016
18	198840.3.dec	1741910H1	566	797
18	198840.3.dec	g1486798	570	979
18	198840.3.dec	969236H1	573	855
18	198840.3.dec	1322802H1	583	918
18	198840.3.dec	5674974H1	593	848
18	198840.3.dec	g4113601	595	974
18	198840.3.dec	g3674968	524	975
18	198840.3.dec	1913281H1	869	1056
18	198840.3.dec	g2837605	872	1052
18	198840.3.dec	533989H1	873	971
18	198840.3.dec	g1384851	836	983
18	198840.3.dec	2761374H1	847	1060
18	198840.3.dec	g677992	672	975
18	198840.3.dec	g1136907	688	983
18	198840.3.dec	g1040531	694	962
18	198840.3.dec	g2177843	737	1052
18	198840.3.dec	2937367H1	771	1047
18	198840.3.dec	667891H1	1	267
18	198840.3.dec	6154236H1	47	367
18	198840.3.dec	g4265077	639	974
18	198840.3.dec	3622345H1	649	710
18	198840.3.dec	g794629	806	983
18	198840.3.dec	1457869H1	818	1056
18	198840.3.dec	g670354	821	1052
18	198840.3.dec	1291302H1	829	1052
18	198840.3.dec	5880838H1	522	789
18	198840.3.dec	5883036H1	522	614
18	198840.3.dec	5881876H1	523	754
18	198840.3.dec	g4686131	524	979
18	198840.3.dec	4784050H1	510	756
18	198840.3.dec	g4833681	521	978
18	198840.3.dec	5882937H1	522	797
18	198840.3.dec	g556213	128	496
18	198840.3.dec	g643590	133	1228
18	198840.3.dec	5020606H1	157	433
18	198840.3.dec	132225H1	210	375
18	198840.3.dec	1715481T7	470	1026
18	198840.3.dec	5000186H2	492	752
18	198840.3.dec	4719977H1	88	351
18	198840.3.dec	2588384H1	598	849
18	198840.3.dec	g1471573	608	971
18	198840.3.dec	g1390212	639	1053
18	198840.3.dec	g1893732	634	978
19	082154.5.dec	g2904866	538	806
19	082154.5.dec	5991508H1	1	273
19	082154.5.dec	5512955H1	1	277
19	082154.5.dec	5512955F6	1	456

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
19	082154.5.dec	531353R6	236	593
19	082154.5.dec	2449285H1	356	594
20	368396.5.dec	g3801673	1	463
20	368396.5.dec	g5395804	1	469
20	368396.5.dec	g2818234	1	488
20	368396.5.dec	2827120H1	1	254
20	368396.5.dec	3295468H1	6	251
20	368396.5.dec	5585616H1	190	412
20	368396.5.dec	g2819673	278	415
20	368396.5.dec	3898288H1	341	610
20	368396.5.dec	388965H1	516	753
20	368396.5.dec	388965R6	517	893
20	368396.5.dec	386150H1	517	787
20	368396.5.dec	568049H1	625	881
20	368396.5.dec	5305447H1	676	951
20	368396.5.dec	840982H1	745	971
20	368396.5.dec	840982R1	745	1335
20	368396.5.dec	4761219H1	912	1196
20	368396.5.dec	5980174H1	942	1170
20	368396.5.dec	4245533H1	1075	1316
20	368396.5.dec	4245533F6	1075	1556
20	368396.5.dec	4245533T6	1076	1517
20	368396.5.dec	5543782H1	1075	1298
20	368396.5.dec	g2884672	1084	1428
20	368396.5.dec	3895077H1	1114	1393
20	368396.5.dec	4904654H2	1131	1400
20	368396.5.dec	1377847H1	1272	1511
20	368396.5.dec	1377895H1	1272	1512
20	368396.5.dec	3186570H1	1285	1543
20	368396.5.dec	5604406H1	1294	1575
20	368396.5.dec	2206040H1	1304	1551
20	368396.5.dec	5507836H1	1341	1562
20	368396.5.dec	3321275H1	1419	1694
20	368396.5.dec	4737695H1	1502	1750
20	368396.5.dec	g5634213	1503	1935
20	368396.5.dec	g5036229	1509	1924
20	368396.5.dec	1517189T6	1526	1932
20	368396.5.dec	3489604H1	1530	1809
20	368396.5.dec	476400H1	1561	1827
20	368396.5.dec	376804H1	1594	1835
20	368396.5.dec	388965T6	1682	2204
20	368396.5.dec	6158381H1	1761	2037
20	368396.5.dec	5597672H1	1767	2029
20	368396.5.dec	5346995H1	1877	2076
20	368396.5.dec	g3146868	1917	2323
20	368396.5.dec	g4268095	1925	2320
20	368396.5.dec	g2913803	2091	2349
20	368396.5.dec	6157520H1	2179	2296
20	368396.5.dec	4180926H1	2225	2492
20	368396.5.dec	6264422H1	2278	2766

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
20	368396.5.dec	4063588H1	2421	2659
20	368396.5.dec	040407H1	2457	2716
20	368396.5.dec	g2795909	2464	4447
20	368396.5.dec	g709369	2464	2783
20	368396.5.dec	g694288	2477	2841
20	368396.5.dec	g766394	2475	2770
20	368396.5.dec	g2787933	3045	3264
20	368396.5.dec	g1267085	3843	4149
20	368396.5.dec	g709370	4105	4447
20	368396.5.dec	g795730	4300	4460
21	349415.4.dec	1471808T6	3019	3286
21	349415.4.dec	1471808H1	3019	3223
21	349415.4.dec	1471808R6	3019	3411
21	349415.4.dec	4552537H1	3278	3514
21	349415.4.dec	859127T6	3424	3933
21	349415.4.dec	2113564T6	3432	3939
21	349415.4.dec	g3181534	3543	3975
21	349415.4.dec	g3804642	3555	3978
21	349415.4.dec	4933708H1	3600	3742
21	349415.4.dec	862833H1	3840	3978
21	349415.4.dec	3074415T6	3847	3974
21	349415.4.dec	g468825	1	4204
21	349415.4.dec	g533522	202	4072
21	349415.4.dec	2113564H1	462	718
21	349415.4.dec	5670744H1	677	844
21	349415.4.dec	g1125015	2400	3418
21	349415.4.dec	g499121	2465	3409
21	349415.4.dec	6246530H1	2798	2928
22	474778.3.dec	3028810H1	859	1044
22	474778.3.dec	818800H1	277	556
22	474778.3.dec	6164205H1	326	657
22	474778.3.dec	6164005H1	327	672
22	474778.3.dec	g4137809	508	953
22	474778.3.dec	3229375H1	2	267
22	474778.3.dec	1955494H1	2	201
22	474778.3.dec	g5446507	196	659
22	474778.3.dec	2431871H1	1	235
23	330933.5.dec	g766379	1791	2088
23	330933.5.dec	1809312T6	1787	2329
23	330933.5.dec	001808H1	2050	2413
23	330933.5.dec	g2107812	2061	2276
23	330933.5.dec	g5369874	2104	2517
23	330933.5.dec	3813723H1	2408	2712
23	330933.5.dec	5901494H1	2416	2718
23	330933.5.dec	1301085F6	2428	2830
23	330933.5.dec	5901773H1	2448	2718
23	330933.5.dec	5813401H1	1873	2202
23	330933.5.dec	g5231675	1818	2273
23	330933.5.dec	g2411104	1820	2269
23	330933.5.dec	g1613942	1940	2275

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
23	330933.5.dec	g2717074	1911	2265
23	330933.5.dec	g1189991	1914	2278
23	330933.5.dec	g4665381	1915	2373
23	330933.5.dec	5813077H1	1916	2202
23	330933.5.dec	g4333942	1918	2276
23	330933.5.dec	g2159566	1911	2287
23	330933.5.dec	5820737H1	1913	2202
23	330933.5.dec	g697258	1736	2083
23	330933.5.dec	g900697	1750	2088
23	330933.5.dec	5210011H1	1759	2027
23	330933.5.dec	2198429T6	1774	2325
23	330933.5.dec	3927974H2	1	160
23	330933.5.dec	5839714H1	1	260
23	330933.5.dec	g1812194	3	308
23	330933.5.dec	4137530H1	4	312
23	330933.5.dec	495251H1	6	171
23	330933.5.dec	495254R6	7	271
23	330933.5.dec	g1615840	8	394
23	330933.5.dec	2602191F6	12	534
23	330933.5.dec	2602191H1	12	303
23	330933.5.dec	493158H1	37	269
23	330933.5.dec	5674628H1	52	320
23	330933.5.dec	5185412H1	61	209
23	330933.5.dec	3297904H1	76	329
23	330933.5.dec	5867451H1	129	261
23	330933.5.dec	5867483H1	130	261
23	330933.5.dec	g714746	139	462
23	330933.5.dec	g1985631	175	569
23	330933.5.dec	191521H1	184	378
23	330933.5.dec	g1198777	206	510
23	330933.5.dec	5296614H1	221	484
23	330933.5.dec	6264549H1	221	641
23	330933.5.dec	3528325H1	347	637
23	330933.5.dec	g1442579	347	683
23	330933.5.dec	5924396H1	397	647
23	330933.5.dec	3054803H1	473	775
23	330933.5.dec	2910408H1	488	739
23	330933.5.dec	266409H1	519	911
23	330933.5.dec	266409R1	520	969
23	330933.5.dec	2289394R6	525	1015
23	330933.5.dec	2289394H1	525	697
23	330933.5.dec	492226R6	575	981
23	330933.5.dec	492226H1	575	833
23	330933.5.dec	3332086H1	626	870
23	330933.5.dec	5163228H1	692	945
23	330933.5.dec	g2159686	777	1093
23	330933.5.dec	2198429F6	824	1252
23	330933.5.dec	2198429H1	824	1074
23	330933.5.dec	3052435H1	841	1121
23	330933.5.dec	g2107811	990	1411

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
23	330933.5.dec	872721H1	1023	1277
23	330933.5.dec	4381555H2	1049	1321
23	330933.5.dec	2837978H1	1050	1297
23	330933.5.dec	2837978F6	1050	1570
23	330933.5.dec	3467189H1	1080	1323
23	330933.5.dec	2280035H1	1110	1373
23	330933.5.dec	g1186783	1111	1283
23	330933.5.dec	998207H1	1160	1423
23	330933.5.dec	3110065H1	1192	1485
23	330933.5.dec	1267448F1	1203	1611
23	330933.5.dec	1267448F6	1203	1750
23	330933.5.dec	1267448H1	1204	1442
23	330933.5.dec	900107H1	1224	1535
23	330933.5.dec	900107R1	1224	1726
23	330933.5.dec	g767718	1254	1925
23	330933.5.dec	3157251H1	1318	1607
23	330933.5.dec	771694H1	1341	1556
23	330933.5.dec	771694R1	1341	1897
23	330933.5.dec	1296960H1	1344	1644
23	330933.5.dec	g1740525	1360	1718
23	330933.5.dec	2848360F6	1362	1812
23	330933.5.dec	2848360H1	1362	1700
23	330933.5.dec	618579H1	1391	1625
23	330933.5.dec	2509950H1	1395	1707
23	330933.5.dec	4342862H1	1428	1778
23	330933.5.dec	5469096H1	1450	1712
23	330933.5.dec	5605680H1	1458	1685
23	330933.5.dec	3222642H1	1481	1787
23	330933.5.dec	g316443	1482	1755
23	330933.5.dec	5691913H1	1503	1805
23	330933.5.dec	5948153H1	1506	1807
23	330933.5.dec	5598313H1	1515	1778
23	330933.5.dec	g5397192	1519	1968
23	330933.5.dec	3578326H1	1536	1792
23	330933.5.dec	3236629H1	1536	1732
23	330933.5.dec	4111759H1	1580	1834
23	330933.5.dec	5821608H1	1950	2202
23	330933.5.dec	g4898000	1960	2374
23	330933.5.dec	1499114H1	1972	2230
23	330933.5.dec	5013907H1	1973	2251
23	330933.5.dec	1267448T6	1982	2478
23	330933.5.dec	g4268930	1984	2273
23	330933.5.dec	g5178020	2005	2285
23	330933.5.dec	g2432366	1884	2275
23	330933.5.dec	5822549H1	1895	2202
23	330933.5.dec	5821996H1	1896	2202
23	330933.5.dec	5817448H1	1898	2202
23	330933.5.dec	3720963H1	1899	2212
23	330933.5.dec	5819913H1	1900	2202
23	330933.5.dec	g1812081	1908	2275

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
23	330933.5.dec	2487759T6	1910	2483
23	330933.5.dec	2653278H1	1630	1869
23	330933.5.dec	g3917068	1878	2287
23	330933.5.dec	g1740526	1884	2277
23	330933.5.dec	266409F1	1839	2273
23	330933.5.dec	4328132H1	1849	2115
23	330933.5.dec	306108H1	1864	2216
23	330933.5.dec	g2000838	1871	2103
23	330933.5.dec	g5656759	1871	2278
23	330933.5.dec	2866509H1	3116	3215
23	330933.5.dec	g3756034	3222	3615
23	330933.5.dec	g953954	3232	3576
23	330933.5.dec	2612962F6	3254	3737
23	330933.5.dec	g1061429	3334	3547
23	330933.5.dec	463141H1	3409	3608
23	330933.5.dec	5015528H1	3416	3693
23	330933.5.dec	g1984806	3531	3896
23	330933.5.dec	2612962H1	3631	3737
23	330933.5.dec	4602495H1	3795	3925
23	330933.5.dec	5786860H1	1968	2275
23	330933.5.dec	707952H1	1970	2218
23	330933.5.dec	1809312H1	1597	1849
23	330933.5.dec	1809312F6	1597	2044
23	330933.5.dec	495254T6	1601	2236
23	330933.5.dec	6559994H1	1613	2178
23	330933.5.dec	6550957H1	1613	2200
23	330933.5.dec	5619308H1	1624	1896
23	330933.5.dec	963906H1	2186	2264
23	330933.5.dec	2042448H1	2203	2273
23	330933.5.dec	1840195H1	2208	2468
23	330933.5.dec	g4334376	2244	2644
23	330933.5.dec	5586733H1	2257	2473
23	330933.5.dec	492226T6	1796	2222
23	330933.5.dec	601199H1	1795	2035
23	330933.5.dec	477137H1	1798	2052
23	330933.5.dec	g2445135	1804	2273
23	330933.5.dec	g2946741	1811	2284
23	330933.5.dec	3972989H1	1817	2077
23	330933.5.dec	6045101J1	2133	2668
23	330933.5.dec	g2336490	2178	2367
23	330933.5.dec	1832323H1	2855	3080
23	330933.5.dec	g5365612	2890	3025
23	330933.5.dec	2487759H1	3036	3261
23	330933.5.dec	g2000839	2812	3146
23	330933.5.dec	5903656H1	2453	2718
23	330933.5.dec	g959363	2529	2823
23	330933.5.dec	1301085H1	2582	2830
23	330933.5.dec	2155586H1	2611	2807
23	330933.5.dec	g803672	2626	2784
23	330933.5.dec	2487759F6	2777	3261

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
23	330933.5.dec	3881086H1	1699	1966
23	330933.5.dec	g900413	1712	2083
23	330933.5.dec	463141T6	3042	3577
23	330933.5.dec	3980529H1	3814	3925
23	330933.5.dec	5568862H1	1648	1917
23	330933.5.dec	1303090H1	1650	1848
23	330933.5.dec	4620879H1	1652	1791
23	330933.5.dec	2327576H1	1663	1924
23	330933.5.dec	5613964H1	1968	2017
23	330933.5.dec	5793991H1	1968	2254
23	330933.5.dec	2581254H1	1591	1835
23	330933.5.dec	g4311841	1751	2206
23	330933.5.dec	761800H1	1752	2038
23	330933.5.dec	1678609H1	1752	1964
23	330933.5.dec	2289394T6	2298	2744
23	330933.5.dec	g3203353	2322	2638
24	998036.2.dec	1389466H1	1	184
24	998036.2.dec	1389427H1	1	173
24	998036.2.dec	799829H1	123	355
24	998036.2.dec	4700320H1	206	477
24	998036.2.dec	5444070H1	222	489
24	998036.2.dec	g4136758	865	1296
24	998036.2.dec	g2583504	932	1304
24	998036.2.dec	4524035H1	945	1209
24	998036.2.dec	4384305H1	842	984
24	998036.2.dec	4386159H1	842	1093
24	998036.2.dec	2915642H1	952	1237
24	998036.2.dec	2915616H1	952	1157
24	998036.2.dec	961104H1	955	1122
24	998036.2.dec	5843637H1	983	1211
24	998036.2.dec	2343721H1	993	1250
24	998036.2.dec	g3754162	1014	1457
24	998036.2.dec	896617H1	1061	1245
24	998036.2.dec	g4114679	856	1293
24	998036.2.dec	904525R6	236	649
24	998036.2.dec	904525H1	236	510
24	998036.2.dec	5610611H1	298	567
24	998036.2.dec	1969343R6	347	745
24	998036.2.dec	1969343H1	347	598
24	998036.2.dec	1603990H1	351	576
24	998036.2.dec	1603974H1	351	580
24	998036.2.dec	4772331H1	365	632
24	998036.2.dec	4994912H1	379	624
24	998036.2.dec	2438688H1	457	680
24	998036.2.dec	4383640H1	498	753
24	998036.2.dec	5906450H1	531	802
24	998036.2.dec	1350441H1	653	890
24	998036.2.dec	g3162658	666	1063
24	998036.2.dec	5623485H1	676	857
24	998036.2.dec	1969343T6	695	1254

TABLE 4

SEQ ID NO:	Template ID	Component ID	Start	Stop
24	998036.2.dec	3723695H1	765	949
25	999304.1.dec	2327457T6	1	364
25	999304.1.dec	2327449H1	4	248
25	999304.1.dec	2327457R6	13	402
25	999304.1.dec	6537441H1	147	499
25	999304.1.dec	5108773H1	196	254

TABLE 5

SEQ ID NO:	Template ID	Tissue Distribution
1	348736.2.oct	Cardiovascular System - 32%, Exocrine Glands - 29%, Hemic and Immune System - 29%
2	025119.6.oct	Unclassified/Mixed - 37%, Germ Cells - 31%
3	474539.1.oct	Embryonic Structures - 44%, Hemic and Immune System - 26%, Male Genitalia - 11%, Digestive System - 11%
4	197170.1.oct	Unclassified/Mixed - 48%, Pancreas - 10%, Digestive System - 10%
5	345638.1.oct	Liver - 17%
6	408784.1.dec	Hemic and Immune System - 57%, Female Genitalia - 21%, Male Genitalia - 14%
7	246526.2.dec	Germ Cells - 11%
8	200488.5.dec	Endocrine System - 100%
10	335916.2.dec	Male Genitalia - 44%, Cardiovascular System - 25%, Exocrine Glands - 25%
11	040422.12.dec	Urinary Tract - 100%
12	977651.2.dec	widely distributed
14	059263.6.dec	Hemic and Immune System - 69%, Respiratory System - 23%
15	196774.3.dec	Hemic and Immune System - 100%
16	233624.11.dec	Digestive System - 100%
17	228585.3.dec	Nervous System - 34%, Germ Cells - 11%
19	082154.5.dec	Cardiovascular System - 33%, Nervous System - 25%, Female Genitalia - 25%
20	368396.5.dec	Unclassified/Mixed - 28%, Hemic and Immune System - 23%
21	349415.4.dec	Skin - 28%, Musculoskeletal System - 25%, Exocrine Glands - 13%, Hemic and Immune System - 13%
23	330933.5.dec	Digestive System - 100%
24	998036.2.dec	Exocrine Glands - 25%, Hemic and Immune System - 25%, Nervous System - 24%
25	999304.1.dec	Digestive System - 50%, Female Genitalia - 30%, Male Genitalia - 20%

TABLE 6

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	PE Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	PE Biosystems, Foster City, CA;	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	PE Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less; Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6; Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less; Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

WO 01/23538 PCT/US00/26085

TABLE 6

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10;1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

CLAIMS

What is claimed is:

1. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25,

b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25,

c) a polynucleotide sequence complementary to a),

d) a polynucleotide sequence complementary to b), and

e) an RNA equivalent of a) through d).

2. An isolated polynucleotide of claim 1, comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:1-25.

3. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 1.

4. A composition for the detection of expression of disease detection and treatment molecule polynucleotides comprising at least one of the polynucleotides of claim 1 and a detectable label.

5. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 1, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

6. A method for detecting a target polynucleotide in a sample, said target polynucleotide comprising a sequence of a polynucleotide of claim 1, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

7. A method of claim 5, wherein the probe comprises at least 30 contiguous nucleotides.

8. A method of claim 5, wherein the probe comprises at least 60 contiguous nucleotides.

9. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 1.

10. A cell transformed with a recombinant polynucleotide of claim 9.

11. A transgenic organism comprising a recombinant polynucleotide of claim 9.

12. A method for producing a disease detection and treatment molecule polypeptide, the method comprising:

a) culturing a cell under conditions suitable for expression of the disease detection and treatment molecule polypeptide, wherein said cell is transformed with a recombinant polynucleotide of claim 9, and

b) recovering the disease detection and treatment molecule polypeptide so expressed.

13. A purified disease detection and treatment molecule polypeptide encoded by at least one of the polynucleotides of claim 2.

14. An isolated antibody which specifically binds to a disease detection and treatment molecule polypeptide of claim 13.

15. A method of identifying a test compound which specifically binds to the disease detection and treatment molecule polypeptide of claim 13, the method comprising the steps of:

a) providing a test compound;

b) combining the disease detection and treatment molecule polypeptide with the test compound for a sufficient time and under suitable conditions for binding; and

c) detecting binding of the disease detection and treatment molecule polypeptide to the test compound, thereby identifying the test compound which specifically binds the disease detection and treatment molecule polypeptide.

16. A microarray wherein at least one element of the microarray is a polynucleotide of claim 3.

17. A method for generating a transcript image of a sample which contains polynucleotides, the method comprising the steps of:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 16 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

18. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence of claim 1, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

19. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 1 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 1 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 April 2001 (05.04.2001)

PCT

(10) International Publication Number
WO 01/23538 A2

(51) International Patent Classification[7]: C12N 15/00, 15/63, C07K 14/47, 16/00, C12Q 1/68, G01N 33/68

(21) International Application Number: PCT/US00/26085

(22) International Filing Date:
22 September 2000 (22.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/156,565	28 September 1999 (28.09.1999)	US
60/168,197	30 November 1999 (30.11.1999)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/156,565 (CIP)
Filed on	28 September 1999 (28.09.1999)
US	60/168,197 (CIP)
Filed on	30 November 1999 (30.11.1999)

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]: 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HODGSON, David, M.** [US/US]; 567 Addison Avenue, Palo Alto, CA 94301 (US). **LINCOLN, Stephen, E.** [US/US]; 725 Sapphire Street, Redwood City, CA 94061 (US). **RUSSO, Frank, D.** [US/US]; 1583 Courdillaeras Road, Redwood City, CA 94062 (US). **SPIRO, Peter, A.** [US/US]; 3875 Park Boulevard, Apt. B16, Palo Alto, CA 94306 (US). **BANVILLE, Steven, C.** [US/US]; 604 San Diego Avenue, Sunnyvale, CA 94086 (US). **BRATCHER, Shawn, R.** [US/US]; 550 Ortega Avenue, #B321, Mountain View, CA 94040 (US). **DUFOUR, Gerard, E.** [US/US]; 5327 Greenridge Road, Castro Valley, CA 94552-2619 (US). **COHEN, Howard, J.** [US/US]; 3272 Cowper Street, Palo Alto, CA 94306-3004 (US). **ROSEN, Bruce, H.** [US/US]; 177 Hanna Way, Menlo Park, CA 94025 (US). **SHAH, Purvi** [IN/US]; 859 Salt Lake Drive, San Jose, CA 95133 (US). **CHALUP, Michael, S.** [US/US]; 183 Alcanes Drive, Apt. 6, Sunnyvale, CA 94086 (US). **HILLMAN, Jennifer, L.** [US/US]; 230 Monrow Drive, #17, Mountain View, CA 94040 (US). **JONES, Anissa, L.** [US/US]; 445 South 15th Street, San Jose, CA 95112 (US). **YU, Jimmy, Y.** [US/US]; 37330 Portico Terrace, Fremont, CA 94536-7901 (US). **GREENAWALT, Lila, B.** [US/US]; 1596 Ballantree Way, San Jose, CA 95118-2106 (US). **PANZER, Scott, R.** [US/US]; 965 East El Camino, #621, Sunnyvale, CA 94087 (US). **ROSEBERRY, Ann, M.** [US/US]; 725 Sapphire Street, Redwood City, CA 94061 (US). **WRIGHT, Rachel, J.** [NE/US]; 339 Anna Way, Mountain View, CA 94043 (US). **CHEN, Wensheng** [CH/US]; 210 Easy Street, #25, Mountain View, CA 94043 (US). **LIU, Tommy, F.** [US/US]; 201 Ottilia Street, Daly City, CA 94014 (US). **YAP, Pierre, E.** [US/US]; 201 Happy Hollow Court, Lafayette, CA 94549-6243 (US). **STOCKDREHER, Theresa, K.** [US/US]; 1596 Ontario Drive, #2, Sunnyvale, CA 94087 (US). **AMSHEY, Stefan** [US/US]; 1541 Canna Court, Mountain View, CA 94043 (US). **FONG, Willy, T.** [US/US]; 572 Cambridge Street, San Francisco, CA 94134 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette

(54) Title: MOLECULES FOR DISEASE DETECTION AND TREATMENT

(57) Abstract: The present invention provides purified disease detection and treatment molecule polynucleotides (mddt). Also encompassed are the polypeptides (MDDT) encoded by mddt. The invention also provides for the use of mddt, or complements, oligonucleotides, or fragments thereof in diagnostic assays. The invention further provides for vectors and host cells containing mddt for the expression of MDDT. The invention additionally provides for the use of isolated and purified MDDT to induce antibodies and to screen libraries of compounds and the use of anti-MDDT antibodies in diagnostic assays. Also provided are microarrays containing mddt and methods of use.

WO 01/23538 A2

Docket No.: PT-1086 USN

"Express Mail" mailing label number EL 856 149 265 US. I hereby certify that this document and referenced attachments are being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR § 1.10, addressed to: Commissioner for Patents, Box Patent Application, Washington, D.C. 20231 on 27 March 2002

By: Nancy Ramos Printed: NANCY RAMOS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Hodgson et al.

Title: MOLECULES FOR DISEASE DETECTION AND TREATMENT

PCT Serial No.: PCT/US00/26085 International Filing Date: 22 September 2000

Examiner: To Be Assigned Group Art Unit: To Be Assigned

Commissioner for Patents
Box Patent Application
Washington, D.C. 20231

SUBMISSION UNDER 37 CFR § 1.821-1.825 SEQUENCE LISTING

Sir:

In accordance with the requirements of 37 CFR § 1.821-1.825, Applicants hereby submit one (1) diskette(s) containing the computer-readable information for the Sequence Listing of the above-identified application. The content of the Sequence Listing paper copy is identical to the computer-readable copy filed with the US Receiving Office. The USPTO is authorized to add whatever is necessary to update the CRF with the current application information.

Respectfully submitted,

INCYTE GENOMICS, INC.

Date: 27 March 2002

Diana Hamlet-Cox
Reg. No. 33,302
Direct Dial Telephone: (650) 845-4639

3160 Porter Drive
Palo Alto, California 94304
Phone: (650) 855-0555
Fax: (650) 845-4166

SEQUENCE LISTING

<110> INCYTE GENOMICS, INC.
HODGSON, David M.
LINCOLN, Stephen E.
RUSSO, Frank D.
SPIRO, Peter A.
BANVILLE, Steven C.
BRATCHER, Shawn R.
DUFOUR, Gerard E.
COHEN, Howard J.
ROSEN, Bruce H.
SHAH, Purvi
CHALUP, Michael S.
HILLMAN, Jennifer L.
JONES, Anissa L.
YU, Jimmy Y.
GREENAWALT, Lila B.
PANZER, Scott R.
ROSEBERRY, Ann M.
WRIGHT, Rachel J.
CHEN, Wensheng
LIU, Tommy F.
YAP, Pierre E.
STOCKDREHER, Theresa K.
AMSHEY, Stefan
FONG, Willy T.

<120> MOLECULES FOR DISEASE DETECTION AND TREATMENT

<130> PT-1086 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/156,565; 60/168,197

<151> 1999-09-28; 1999-11-30

<160> 25
<170> PERL Program

<210> 1
<211> 569
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 348736.2.oct

<400> 1
ggggatggaa ccccttatct caggatcaca gcatagatat ctatttgttt atctgcctct 60
tgttgctgat ggctatgtgt catgatcagt ggtcccatga ggacacagtg gcactagtgg 120
gcaaagtttc tcctctgagt tggggatcct gggagaagga ggatgggcaa gggtagatat 180
gttggggaag tggattgtgg gtcttcagca cgggccagca cgggccattt atggtttcat 240
ctatccccat cataacagag cagtgtgacc tttgaagacg tggctgtaaa cttttccctg 300
gaggaatgga gtcttcttaa tgaggctcag ggatgcctgt accatgatgt gatgctggag 360
accttgacac ttatatcctc cctggtaagg tactcatact taactgtgac ctgagttagt 420
ctctgccccc ccccttttatt cctcttggta ataacgtctt tctcacatca ggactgtggc 480
acagcttcat tctccaattc tctggatcag ttctgtggtt ggtagtactg agatacatgt 540
actgccctct cctttccttg agcagcccc 569

<210> 2
<211> 673
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 025119.6.oct

<220>
<221> unsure
<222> 6-7, 42, 46, 162
<223> a, t, c, g, or other

<400> 2
cttggnnccc gggccgggga ggctttctcg ggcgcaggag gntccncagg cccaggccag 60
gccaggggag gcagccgatc cgtcgtcggg gttgacagtt accatggcgc cgcctctggc 120
cccgctccct ccccgggacc caaacggggc cggacccgag tngaggaagc ccgggactgt 180
gagcttcgcg gacgtggccg tgtacttctc cccggaggag tggggctgtc tgcggcccgc 240
gcagagggct ctgtaccggg acgtgatgca ggagacctac ggccacctgg gcgcgctcgg 300
attcccaggc cccaaaccag ccctcatctc ttggatggaa caggagagtg aggcttggag 360
ccccgccgcc caggatcctg agaaggggga aagactggga ggagctcgga gaggagatgt 420
cccaaacagg aaggaagagg aaccggagga agtcccaaga gccaaagggc ctagaaaggc 480
tcctgtgaag gagagtcctg aagtgctggt ggaacgcaac cctgacccag ctattagcgt 540
ggccccggca cgggcacagc cacccaaaaa tgctgcctgg gacccgacca caggagcaca 600
gcccccggca cccataccca gcatggatgc tcaggccggc cagcggcgcc acgtgtgcac 660
ggactgcggc cgc 673

<210> 3
<211> 429
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 474539.1.oct

<220>
<221> unsure
<222> 416, 424
<223> a, t, c, g, or other

<400> 3
gggcgtgctc agcaaataca ccaacctcct ccagggctgg cagaacaggt acttcgtact 60
ggatttcgag gctggcatcc tgcagtattt tgtgaatgag caaagcaaac accagaagcc 120
tcgaggagtc ctgtctttat ctggagccat agtgtccctg agcgatgaag ctccccacat 180
gctggtggtg tactctgcta atggagagat gtttaaactg agagctgctg atgcaaaaga 240
gaaacaattc tgggtgactc agcttcgagc ttgtgccaaa taccacatgg aaatgaattc 300
taagagtgct ccaagctccc gaagccgaag tctcactttg ctcccacatg gaacacccaa 360
ttctgcgtct ccctgtagcc agagacacct cattgttggg ggcccccggt gttgtnacaa 420
tcangcatc 429

<210> 4
<211> 1517
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 197170.1.oct

<220>
<221> unsure
<222> 1428, 1440, 1452, 1498
<223> a, t, c, g, or other

<400> 4
tcccacctcc gcggccctct ccttgcttcc ccccccaccg ccggccaaga aggccaagct 60
gaaggccgcg ggtatggcca gcccctgggg gaagcaggac ctctcggccg ccgcagccgc 120
cggcattttc tgggcctctg atgtggagcc gtctcctctc aacctctcct caggcccaga 180

gccagcacga gacatccgct gcgagttctg tggtgagttc ttcgagaacc gcaagggcct 240
ctcgagccac gcgcgctccc atctgcggca aatgggcgtg accgagtggt acgtcaatgg 300
ctcgcccatc gacacgctgc gggagatcct gaagagacgg acccagtctc ggcctggtgg 360
acctcccaac ccaccagggc caagcccaaa agccctggcc aagatgatgg gcggcgcagg 420
tcctggcagc tcactggaag cccgcagccc ctcggacctt cacatctcac ccttggccaa 480
gaagttgcca ccaccaccgg gcagcccccc gggccactca ccaactgcct ctcctcctcc 540
tacggcccga aagatgttcc caggcctggc tgcaccctcc ttgcccaaga agctgaagcc 600
tgaacaaata cgggtggaga tcaagcggga gatgctgccg ggggcccttc atggggaact 660
gcacccatct gagggtccct ggggggcacc acgggaagac atgacacccc tgaacctgtc 720
gtcccgggca gagccggtgc gggacatccg ctgtgagttc tgcggcgagt tcttcgagaa 780
ccgcaagggc ctgtcgagtc acgcgcgctc acacctgcgg cagatgggtg tgaccgagtg 840
gtccgtcaat ggttcgccca tcgacacact gcgagagatc ctcaagaaga agtccaagcc 900
gtgcctcatc aagaaggagc caccggctgg agacctggcc cctgccctgg ctgaggacgg 960
gcctcccacc gtggcccctg ggcccgtgca gtccccactg ccgctgtcgc ccctggctgg 1020
ccggccaggc aaaccaggtg caggggccgg cccaggttcc tcgtgagctc agcctgacgc 1080
ccatcactgg ggccaagccc tcagccactg gctacctggg ctcagtggca gccaagcggc 1140
ccctgcagga ggaccgcctc ctcccagcag aggtcaaggc caagacctac atccagactg 1200
aactgccctt caaggcaaag acccttcatg agaagacctc ccactcctcc accgaggcct 1260
gctgcgagct gtgtggcctt tatttgaaaa ccgcaaggcc ctggccagcc acgcacgggc 1320
acacctgcgg cagttcggcg tgaccgagtg gtgcgtcaat ggctcgccca tcgagacact 1380
gagcgagtgg atcaaacacc ggccccagaa ggtgggcgcc taccgcanct acatccaggn 1440
cggccgccct tnaccaagaa gttcgcagtg ccggccatgg ccgtgacagt gacaagcngc 1500
cgtccctggg gctggca 1517

<210> 5
<211> 2185
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 345638.1.oct

<220>
<221> unsure
<222> 2153, 2159, 2175
<223> a, t, c, g, or other

<400> 5
ggaggaggag gtggggctgg cgctgaagcc ggatccggat ccggtgctgt gcacactggt 60
gggggagagt ccgacgcgcc tggctaggag cgccgaccgc agggcctcta cggaccttac 120
tagaaaaatg aaacctgatg aaactcctat gtttgaccca agtctactca aagaagtgga 180
ctggagtcag aatacagcta cattttctcc agccatttcc ccaacacatc ctggagaagg 240
cttggttttg aggcctcttt gtactgctga cttaaataga ggtttttta aggtattggg 300
tcagctaaca gagactggag ttgtcagccc tgaacaattt atgaaatctt ttgagcatat 360
gaagaaatct ggggattatt atgttacagt tgtagaagat gtgactctag gacagattgt 420
tgctacggca actctgatta tagaacataa attcatccat tcctgtgcta agagaggaag 480
agtagaagat gttgttgtta gtgatgaatg cagaggaaag cagcttggca aattgttatt 540
atcaaccctt actttgctaa gcaagaaact gaactgttac aagattaccc ttgaatgtct 600
accacaaaat gttggtttct ataaaaagtt tggatatact gtatctgaag aaaactacat 660
gtgtcggagg tttctaaagt aaaaatcttg taagaaaatt gtcaaagggg ctaatgctac 720
aaggctacac tcttcctaga gttgaaatat tttgttgctg cagccgagtg acctccataa 780
atactggact gaaaaaacat tgtaatacta caagtataat gacatttaga agattacttt 840
gggctggtgg gacatgctgt gaatttagat tacaaatgaa tattataaag gggatgattt 900
ttaaccaaag gaatatattt ttaacttgaa tcttttcttg cattgtattt ttctaaaagt 960
ttggcttcct ttcttggtag tcaagagtat gggtaataag gagttatatg tctgctatct 1020
gtgttgctca tttaaaaaaa gtatacattg aataaggctg tttatcacat gcataaaatt 1080
aaatattttt gtttcaaaga aacatctcaa tacacttagg ggtgtattgt ttcccacata 1140
ttaagtcagg gtggataaat tagttattat aactaaacat agtatagtcc aacattcgtt 1200
gatcccaata caggcaaaca acctggtcaa ccttttgaag tagaagaaat gaaaattact 1260
tgacaagatt aaaagtaaaa caatttaaat gttttactga aagtttatat agtatagtct 1320
atgtagataa aaagtaccac ttgtcttttc tgtgaattat gactattcat ttgttaaaaa 1380
tacctaagag caattatagt gggacatctt aggtcctctg taaacagtga attagcaaac 1440
ctcagcctat gtgtttctac cctgattttt ttcttttcat gggtatctga agcctctaag 1500
ttttttcaaa aatggagtat cacaaaattg agtgaaacac aatacttaat gtattgtact 1560
agattgccaa attcataaaa tgttaatgga agctttttga tgtgattata atggcactat 1620
tctggtcatt atcctatttt gattttattt aatttttttaa agttgaagaa ttaaatattt 1680

taatggttct aatcttttgc attccatgtt gcattaaacc tgtttatatg agtagtcttc 1740
tgttagaatc acatctgtgc ttttcttgag tctgctgttg aactattaga ttaagtcata 1800
attcataaaa ttttagttta atgtgctctt tgtaaaatga aattgtaaag aaaataccag 1860
tgtttctcat cccattgact cacaccacgt catctggatt ttggatttcc ctccatgcag 1920
ccagctatag ttggctttcc aaaacaacag aaatccttca ccaatagagt gcactactta 1980
cctgcttata gcctatacag acgaactgat ctgtccttcg tgaaacgcaa caaagctagt 2040
tctgtctttt cagaagtcct acaaccttga caaagagtag ttttatcagg taaatcctgg 2100
taattaaaaa cgcatgtttt taaaaattag cctggtaagg ccgggtgcag tgnctcagnc 2160
tcccaaagtg ctggnatgac aggtg 2185

<210> 6
<211> 397
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 408784.1.dec

<400> 6
ccagcccaca aggctccttt tccttttga tccattcaaa aattactcat tgcaaattcc 60
cgaaccatcc ggctcgggct ccttccctgg cgatggctgg ccgctgagcc atggctcagt 120
acggccaccc cagtccgctc ggcatggctg cgagagagga gctgtacagc aaagtcaccc 180
cccggaggaa ccgccaacag cgccccggca ccatcaagca tggatcggcg ctggacgtgc 240
tcctctccat ggggttcccc agagcccgcg cacgaaaagc cttggcatcc acgggaggaa 300
gaagtgttca gacagcatgt gactggttat tctcccatgt cggtgacccc ttcctggatg 360
acccccctgcc ccgggagtac gtcctctacc ctccgtc 397

<210> 7
<211> 2815
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 246526.2.dec

<400> 7
ccgggagagc tcgatgggct tctcctgcgc gccgcccggt gtctggccga gtccagagag 60
ccgcggcgcc tcgttccgag gagccatcgc cgaagcccga ggccgggtcc cgggttgggg 120
actgcagggg aaggcagcgg cggggcggcg ggagccccac cggggtctgg gactggggaa 180
ctgcctccgg cttcacgatg ccagtatgga cagaatagct tatgatgctt atccccaccc 240
accacttccg aaacattgag cggaaaccag aatacctcca gccagagaag tgtgtcccac 300
ccccctaccc tggtcctgtg ggaaccatgt ggtttatccg tgacggctgt ggcatcgcct 360
gtgccatcgt tacctggttt ctggtcctct atgcggagtt cgtggtcctc tttgtcatgc 420
tgattccatc tcgagactac gtgtatagca tcatcaacgg aattgtgttc aacctgctgg 480
ccttcttggc cctggcctcc cactgccggg ccatgctgac ggaccccggg gcagtgccca 540
aaggaaatgc cactaaagaa ttcatcgaga gtttacagtt gaagcctggg caggtggtgt 600
acaagtgccc caaatgctgc agcatcaagc ccgaccgagc ccaccactgc agtgtttgta 660
agcggtgcat tcggaagatg gaccaccact gtccctgggt caacaactgt gtaggcgaga 720
acaaccagaa gtacttcgtc ctgtttacaa tgtacatagc tctcatttcc ttgcacgccc 780
tcatcatggt gggattccac ttcctgcatt gctttgaaga agattggaca aagtgcagct 840
ccttctctcc acccaccaca gtgattctcc ttatcctgct gtgctttgag ggcctgctct 900
tcctcatttt cacatcagtg atgtttggga cccaggtgca ctccatctgc acagatgaga 960
cgggaataga acaattgaaa aaggaagaga gaagatgggc taaaaaaaca aaatggatga 1020
acatgaaagc cgtttttggc caccccttct ctctaggctg ggccagcccc tttgccacgc 1080
cagaccaagg gaaggcagac ccgtaccagt atgtggtctg aaggaccccg accggcatgg 1140
ccactcagac acaagtccac accacagcac taccgtccca tccgttctca tgaatgttta 1200
aatcgaaaaa gcaaaacaac tactcttaaa acttttttta tgtctcaagt aaaatggctg 1260
agcattgcag agaaaaaaaa aagtccccac attttatttt ttaaaaacca tcctttcgat 1320
ttcttttggt gaccgaagct gctctctttt ccttttaaaa tcacttctct ggcctctggt 1380
ttctctctgc tgtctgtctg gcatgactaa tgtagagggc gctgtctcgc gctgtgccca 1440
ttctactaac tgagtgagac atgacgctgt gcgtggatgg aatagtctgg acacctggtg 1500
ggggatgcat gggaaagcca ggagggccct gacctcccac tgcccaggag gcagtggcgg 1560
gctccccgat gggacataaa acctcaccga agatggatgc ttacccctg aggcctgaga 1620
agggcaggat cagaagggac cttggcacag cgacctcatc ccccaagtgg acacggtttg 1680
cctgctaact cgcaaagcaa ttgcctgcct tgtactttat gggcttgggg tgtgtagaat 1740

gattttgcgg gggagtgggg agaaagatga aagaggtctt atttgtattc tgaatcagca 1800
attatattcc ctgtgattat ttggaagagt gtgtaggaaa gacgtttttc cagttcaaaa 1860
tgccttatac aatcaagagg aaaaaaaatt acacaatttc aggcaagcta cgttttcctt 1920
tgtttcatct gcttcctctc tcaccacccc atctccctct cttccccagc aagatgtcaa 1980
ttaagcagtg tgaattctga ctgcaatagg caccagtgcc caacacatac agccccacca 2040
tcatcccctt ctcattttat aaacctcaaa gtggattcac tttctgatag ttaaccccca 2100
taaatgtgca cgtacctgtg tcttatctat attttaacct gggagactgt tgtcctggca 2160
tggagatgac catgatgctg gggttacctc acagtcccca ccctttcaaa gttgacatat 2220
ggccatccca ttggccagaa tccacagaca cacctaagcc tgtggcactg ggacagaata 2280
gattttccat ttgagaggca cttcctgtgt cagtcttgtt tgaaggaggt ggtgatggtg 2340
gatagaggtg aaggaggtag ggagtgccct ccaagtgcaa aaataacaaa tatgattatt 2400
gaccatcggg gaattctcac acattgattt gtttttaag caattgccag aaacccccctt 2460
tttttagctt ttgcttgggg tgggggtagg agttaaggtt tattcaatcc tgtcctgggt 2520
agggcgaaag ttaatctagc catgtgattt ttcagaaaag taagtggaac atgctgccac 2580
ttttcaattc tgtcagtgct tccacatgga aacaaaatgc aataaaattt ttccaaaacc 2640
tgttctgatt tagctctctc ttgaggtgtt acccttagtg ggaggccgac tatccacaat 2700
ctacttgagt tttctctggt tgggtgtttg tttcattgct ctgtctcttg aatgaggata 2760
ctttatttttt tttgttttaa aatgcattta tggtccctct cttgaaccag cttgc 2815

<210> 8
<211> 771
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 200488.5.dec

<220>
<221> unsure
<222> 7
<223> a, t, c, g, or other

<400> 8
ccgagangtg cagcggcaca gctgtcgcgc cagtcgcaac agaagcaggt ccgaggcaca 60
gcccgatccc gccatggagc agccgaggaa ggcggtggta gtgacgggat ttggcccttt 120
tggggaacac accgtgaacg ccagttggat tgcagttcag gagctagaaa agctaggcct 180
tggcgacagc gtggacctgc atgtgtacga gattccggtt gagtaccaaa cagtccagag 240
actcatcccc gccctgtggg agaagcacag tccacagctg gtggtgcatg tgggggtgtc 300
aggcatggcg accacagtca cactggagaa atgtggacac aacaagggct acaaggggct 360
ggacaactgc cgcttttgcc ccggctccca gtgctgcgtg gaggacgggc ctgaaagcat 420
tgactccatc atcgacatgg atgctgtgtg caagcgagtc accacgttgg gcctggatgt 480
gtcggtgacc atctcgcagg atgccggcag gaaaaaaccc ttccctgcca aaggtgactg 540
tgttttctgc cgccgaagga gggcccggtc cctccaggct cagtgtggct tctccctgac 600
ccccgcccta gaacttttgc cagtgccttt tctgaaactc ctgtgtcccg ggccccccag 660
gcggagaagg atatgccgga ttctgcctgg ggctgggctc taggagaccc caaatttgac 720
accacagaaa gcagataaaa cacttgaaat acgcagaaaa aaaaaaaaag g 771

<210> 9
<211> 2431
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 474878.1.dec

<220>
<221> unsure
<222> 2427
<223> a, t, c, g, or other

<400> 9
ctgatttcga gtttccggtc aggttaggcc ggggggggtgc ggtcctggtc ggaaggaggt 60
ggagagtcgg gggtcaccag gcctatcctt ggcgccacag tcggccaccg gggctcgccg 120
ccgtcatgga gagcggaggg cggccctcgc tgtgccagtt catcctcctg ggcaccacct 180
ctgtggtcac cgccgccctg tactccgtgt accggcagaa ggcccgggtc tcccaagagc 240

tcaagggagc taaaaaagtt catttgggtg aagatttaaa gagtattctt tcagaagctc 300
caggaaaatg cgtgccttat gctgttatag aaggagctgt gcggtctgtt aaagaaacgc 360
ttaacagcca gtttgtggaa aactgcaagg gggtaattca gcggctgaca cttcaggagc 420
acaagatggt gtggaatcga accacccacc tttggaatga ttgctcaaag atcattcatc 480
agaggaccaa cacagtgccc tttgacctgg tgccccacga ggatggcgtg gatgtggctg 540
tgcgagtgct gaagcccctg gactcagtgg atctgggtct agagactgtg tatgagaagt 600
tccacccctc gattcagtcc ttcaccgatg tcatcggcca ctacatcagc ggtgagcggc 660
ccaaaggcat ccaagagacc gaggagatgc tgaaggtggg ggccaccctc acaggggttg 720
gcgaactggt cctggacaac aactctgtcc gcctgcagcc gcccaaacaa ggcatgcagt 780
actatctaag cagccaggac ttcgacagcc tgctgcagag gcaggagtcg agcgtcaggc 840
tctggaaggt gctggcgctg gtttttggct ttgccacatg tgccaccctc ttcttcattc 900
tccggaagca gtatctgcag cggcaggagc gcctgcgcct gcaagcagat gcaggaggag 960
ttccaggagc atgaggccca gctgctgagc cgagccaagc ctgaggacag ggagagtctg 1020
aagagcgcct gtgtagtgtg tctgagcagc ttcaagtcct gcgtctttct ggagtgtggg 1080
cacgtttgtt cctgcaccga gtgctaccgc gccttgccag agcccaagaa gtgccctatc 1140
tgcagacagg cgatcacccg ggtgataccc ctgtacaaca gctaatagtt tggaagccgc 1200
acagcttgac ctggaagcac ccctgccccc ttttcaggga tttttatctc gaggcctttg 1260
gaggagcagt ggtgggggta gctgtcacct ccaggtatga ttgagggagg aattgggtag 1320
aaactctcca gacccacgcc tccaatggca ggatgctgcc tttcccacct gagaggggac 1380
cctgtccatg tgcagcctca tcagagcctc accctgggag gatgccgtgg cgtctcctcc 1440
caggagccag atcagtgtga gtgtgactga aaatgcctca tcacttaagc accaaagcca 1500
gtgatcagca gctcttctgt tcctgtgtct tctgtttttt tctggtgaat cgttgcttgc 1560
tgtggacttg gtggaggact cagaggggag gaaaggctgg gccccgagta caacggatgc 1620
cttgggtgct gcctccgaag agactctgcc gcagcttttc ttctttttcc tcatgccccg 1680
ggaaacagtc tttcttcaga attgtcaggc tgggcaggtc aacttgtgtt cctttccccc 1740
cacctgcttg cctccttaac gcctgcacgt gtgtgtagag gacaaaagaa agtgaagtca 1800
gcacatccgc ttctgcccag atggtcgggg ccccgggcaa cagattgaag agagatcatg 1860
tgaagggcag ttggtcaggc aggcctcctg gtttcgccac tggccctgat ttgaactcct 1920
gccacttggg agagctcggg gtggtccctg gttttccctc ctggagaatg aggcgcagag 1980
gcctcgcctc ctgaaggacg cagtgtggat gccactggcc tagtgtcctg gcctcacagc 2040
ttccttgcaa ggctgtcaca aggaaaagca gccggctggc accctgagca tatgccctct 2100
tggggctccc tcatccagcc cgtcgcagct ttgacatctt ggtgtactca tgtcgcttct 2160
ccttgtgtta ccccctccca gtattaccat ttgcccctca cctgcccttg gtgagccttt 2220
tagtgcaaga cagatggggc tgttttcccc cacctctgag tagttggagg tcacatacac 2280
agctcttttt ttattgccct tttctgcctc tgaatgttca tctctcgtcc tcctttgtgc 2340
aggcgaggaa ggggtgccct caggggccga cactagtatg atgcagtgtc cagtgtgaac 2400
agcagaaatt aaacatgttg caaccanaaa a 2431

<210> 10
<211> 2064
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 335916.2.dec

<220>
<221> unsure
<222> 1377, 1387
<223> a, t, c, g, or other

<400> 10
ggactgttga cctgcagtgc tctctgtgga caaagggtgt gagcagctgc tggaagagga 60
ggtgctcctg aagacactgc ggccggcccg cctgtgccct tggtgccctg gctgcctaga 120
gagcctcacc cctgggccct ggggccagga ctccaggact ctgactacct gccctccccc 180
agcctcagcg gctgcacctc ctcgttagta ctgatgcact gacctcggca cacagctggg 240
aggggttggg ggctgggtca tggctgctcc caggccccac ccaggctcct gagcctagaa 300
ggtgaaaaga ggactctcag gggctcacag gggctctcac tgctggttgg ccctgccctc 360
ccttcccccc cagcagggtg cccggaagct ggaaccttgt tatctgggta attagtttca 420
gaccctgcac tgaggccggc caggtctcgg ggctgcctcc cataggttgt gcaccctgac 480
cccgagaggg aggcgaggcg ctgcttgtcg acagctagag gctggcctgg ggagcaggtt 540
tggggtgccc tcccacactg ccctccctgc cccggcccat gccccccagg gctgcctggg 600
cctggttatt gtgtggggcc tcctgaccca gccaagggca cgaagctctg ggaaggggat 660
gcccccgagg gtgccagtcc agctagctgc cccacccctc aggcccagcc tggccccaa 720
gctccccact ctggtgcccc gagcagccct gtgggcaagc agccgccgcc atggccgagc 780
acctggagct gctggcagag atgcccatgg tgggcaggat gagcacacag gagcggctga 840

agcatgccca gaagcggcgc gcccagcagg tgaagatgtg ggcccaggct gagaaggagg 900
cccagggcaa gaagggtcct ggggagcgtc cccggaagga ggcagccagc caagggctcc 960
tgaagcaggt cctcttccct cccagtgttg tccttctgga ggccgctgcc cgaaatgacc 1020
tggaagaagt ccgccagttc cttgggagtg gggtcagccc tgacttggcc aacgaggacg 1080
gcctgacggc cctgcaccag tgctgcattg atgatttccg agagatggtg cagcagctcc 1140
tggaggctgg ggccaacatc aatgcctgtg acagtgagtg ctggacgcct ctgcatgctg 1200
cggccacctg cggccacctg cacctggtgg agctgctcat cgccagtggc gccaatctcc 1260
tggcggtcaa caccgacggg aacatgccct atgacctgtg tgatgatgag cagacgctgg 1320
actgcctgga gactgccatg gccgaccgtg gtaggcatca cccaggacag catcgangcc 1380
gcccggnccg tgccagaact gcgcatgctg gacgacatcc ggagccggct gcaggccgga 1440
gcagacctcc atgccccct ggaccacggg ccacctgtgc acgtccaccc caacgggttc 1500
agcgagcggt ccctgtgtgg aacaccgagc cagcctgagc gctaaggacc aagacggctg 1560
ggagccgctg cacgccgcgg cctactgggg ccaggtgcct ggtggagctg ctcgtggcgc 1620
acgggccga cctgaacgca aagtccctga tggacgagac gccccttgat gtgtgcgggg 1680
acgaggaggt gcgggccaag ctgctggagc tgaagcacaa gcacgacgcc ctcctgcgcg 1740
cccagagccg ccagcgctcc ttgctgcgcc gccgcacctc cagcgccggc agccgcggga 1800
aggtggtgag gcgggtgagc ctaacccagc gcaccgacct gtaccgcaag cagcacgccc 1860
aggaggccat cgtgtggcaa cagccgccgc ccaccagccc ggagccgccc gaggacaacg 1920
atgaccgcca gacaggcgca gagctcaggc cgccgccccc ggaggtgagc gccccgtccc 1980
tgctccgccc agcgcagggg tgggcctggc tctgccctgg ttctctctcc gcttggaccc 2040
cctcctgcct gtgtcaggaa tggt 2064

<210> 11
<211> 1421
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 040422.12.dec

<220>
<221> unsure
<222> 913
<223> a, t, c, g, or other

<400> 11
cttgttggta tgtgtagcgg cagtggccgc cggcggagca gtctgagccc gacgatgagg 60
ccggggacgg gagctgagcg tggaggcctc atggtgagtg aaatggagag ccatcctccc 120
tcgcagggtc ctggggacgg ggagcggaga ttgtccggct caagcctctg ctccggctct 180
tgggtctctg ctgacggctt cctgaggaga cggccctcga tggggcaccc tggcatgcat 240
tatgccccaa tgggaatgca ccctatgggt cagagagcga atatgcctcc tgtacctcat 300
ggaatgatgc cgcagatgat gccccctatg ggagggccac caatgggaca aatgcctgga 360
atgatgtcgt cagtaatgcc tggaatgatg atgtctcata tgtctcaggc ttccatgcag 420
cctgccttac cgccaggagt aaatagtatg gatgtagcag caggtacagc atctggtgca 480
aaatcaatgt ggactgaaca taaatcacct gatggaagga cttactacta caacactgaa 540
accaaacagt ctacctggga gaaaccagat gatcttaaaa cacctgctga gcaactctta 600
tctaaatgcc cctggaagga atacaaatca gattctggaa agccttacta ttataattct 660
caaacaaaag aatctcgctg ggccaaacct aaagaacttg aggatcttga aggataccag 720
aataccattg ttgctggaag tcttattaca aaatcaaacc tgcatgcaat gatcaaagct 780
gaagaaagca gtaagcaaga agagtgcacc acaacatcaa cagccccagt ccctacaaca 840
gaaattccga ccacaatgag caccatggct gctgccgaag cagcagctgc tgttgttgca 900
gcagcagcag cgngcagcac gagcagcagc tgcagccaat gctaatgctt ccacttctgc 960
ttctaatact gtcagtggaa ctgttccagt tgttcctgag cctgaagtta cttccattgt 1020
tgctactgtt gtagataatg agaatacagt aactatttca actgaggaac aagcacaact 1080
tactagtacc cctgctattc aggatcaaag tgtggaagta tccagtaata ctggagaaga 1140
aacatctaag caagaaactg tagctgattt tactcccaaa aaagaagagg aggagagcca 1200
accagcaaag aaaacataca cttggaatac aaaggaagag gcaaagcaag cttttaaaga 1260
attattgaaa gaaaagcggg taccatcgaa tgcttcatgg gagcaggcta tgaaaatgat 1320
tattaatgat ccacgataca gtgctttggc aaagttaagt gaaaaaaagc aagcctttaa 1380
tgcctataaa gtccagacag aaaaagaaaa aaaagggcgg c 1421

<210> 12
<211> 1096
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 977651.2.dec

<400> 12
ggttccccgg cctctcttgg tcagggtgac gcagtagcct gcaaacctcg gcgcgtaggc 60
caccgcactt atccgcagca ggaccgcccg cagccggtag ggtgggctct tcccagtgcc 120
cgcccagcta ccggccagcc tgcggctgcg cagatctttc gtggttctgt caggggagac 180
ccttaggcac tccggactaa gatggcggcg acggccaggg cggggctggg gagctgcggc 240
tgttgccgcc ggggctgcgc aggcggtttc tgtcataagt tgaagaatcc atacaccatt 300
aagaaacagc ctctgcatca gtttgtacaa agaccacttt tcccactacc tgcagccttt 360
tatcacccag tgagatacat gtttattcaa acacaagata ccccaaatcc aaacagctta 420
aagtttatac caggaaaacc agttcttgag acaaggacca tggattttcc caccccagct 480
gcagcatttc gctcccctct ggctaggcag ttatttagga ttgaaggagt aaaaagtgtc 540
ttctttggac cagatttcat cactgtcaca aaggaaaatg aagaattaga ctggaattta 600
ctgaaaccag atatttatgc aacaatcatg gacttctttg catctggctt acccctggtt 660
actgaggaaa caccttcagg agaagcagga tctgaagaag atgatgaagt tgtggcaatg 720
attaaggaat tgttagatac tagaatacgg ccaactgtgc aggaagatgg aggggatgta 780
atctacaaag gctttgaaga tggcattgta cagctgaaac tccagggttc ttgtaccagc 840
tgccctagtt caatcattac tctgaaaaat ggaattcaga acatgctgca gttttatatt 900
ccggaggtag aaggcgtaga acaggttatg gatgatgaat cagatgaaaa agaagcaaac 960
tcaccttaaa ataatctgga ttttctttgg gcataacagt cagacttgtt gataatatat 1020
atcaagtttt tattattaat atgctgagga acttgaagat taataaaata tgctcttcag 1080
agaatgataa aaaaaa 1096

<210> 13
<211> 590
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 012432.5.dec

<400> 13
gggcggaaga ggtgggctgg tggaggcggg gtcgagatgg cggcgccttt gaggattcag 60
agcgactggg cgcaagcctc caggaaggat gaaggggagg cctggctgag ctgtcatccc 120
ccagggaaac catctttgta tggcagcctg acttgtcaag gaattggcct agatggcatc 180
ccagaggtta cagcttcaga aggatttact gtgaatgaaa taaacaagaa aagcattcat 240
atttcatgtc caaaggaaaa tgcatcttct aagtttttgg caccatatac tacttttcc 300
agaattcata caaagagtat aacatgcctg gacatttcca gcagaggagg tcttggtgtg 360
tcttctagta ctgacgggac catgaaaatc tggcaggctt ccaatggaga actcaggaga 420
gtattgggaa ggacatgtgt ttgatgtgaa ttgttgcagg tttttcccat caggccttgt 480
ggtcctgagt gggggaatgg atgcccagct gaagatatgg tcagctgaag atgctagctg 540
cgtggtgacc ttcaaaggtc acaaaggagg tatcctggga tacagccatc 590

<210> 14
<211> 2109
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 059263.6.dec

<400> 14
ccccctactg ctacccacag ggcccccact ccacctgctc ccagacgagg ccaactcctg 60
gccaagctta ggggcacagc ggaggcgctc tgtttctgat ttttctcgcc ttccttggag 120
atgcctgtgc ttggaaggga aggcagaaca ttgcatcttg gaacaaatct gcttttgatc 180
atgcaaatga atgcctggat aatgtaggca gactgtcaat ttcaccagtt agaaagaaag 240
agaaaagggg gagaaattcc ccatgacagc gactgatgaa gaatttcaat agaaagctgc 300
tacttcagaa aataagatca tttgctgcga atggagaaca tctcaggcag ccctgatgct 360
ccaccggctc tgggcatcac cagcggcccc agggaaaaag aaagaaatgg gaaacagcat 420
gaaatccacc cctgcgcctg ccgagaggcc cctgcccaac ccggagggac tggatagcga 480
cttccttggc cgtgctaagt gactacccgt ctcctgacat cagccccccg atattccgcc 540
gaggggagaa actgcgtgtg atttctgatg aaggggggctg gtggaaagct atttctctta 600
gcactggtcg agagagttac atccctggaa tatgtgtggc cagagtttac catggctggc 660

tgtttgaggg cctgggcaga gacaaggccg aggagctgct gcagctgcca gacacaaagg 720
tcggctcctt catgatcaga gagagtgaga ccaagaaagg gttttactca ctgtcggtga 780
gacacaggca ggtaaagcat taccgcattt tccgtctgcc caacaactgg tactacattt 840
ccccgaggct caccttccag tgcctggagg acctggtgaa ccactattct gaggtggctg 900
atggcctgtg ctgtgtgctc accacgccct gcctgacaca aagcacggct gccccagcag 960
tgagggcctc cagctcacct gtcaccttgc gtcagaagac tgtggactgg aggagagtgt 1020
ccagactgca ggaggacccc gagggaacag agaacccgct tggggtagac gagtcccttt 1080
tcagctatgg ccttcgagag agcattgcct cttacctgtc cctgaccagt gaggacaaca 1140
cctcctttga tcgaaagaag aaaagcatct ccctgatgta tggtggcagc aagagaaaga 1200
gctcattctt ctcatcacca ccttactttg aggactagcc aagaacagac acaatggttc 1260
atgcccaaaa ggaacagaag ttccaactat tgcctgggat cttgcgaaaa gcgaggttcc 1320
ctgatccctg ggagcctcac gtattttaga agccaagaga agccacatgg agactcaaat 1380
tcgcatcttc tctatccaca tcatgaccaa aggaacccct ccctggtgtc tgatcagggc 1440
tgtggcatca cgaaacattg gatcatgaca tgtcgggcga tgcttggaag agcccagcat 1500
gtatgtatgc acacattgtg tgtgtgggaa ggacaaagcc actctcacaa gaaagggcac 1560
caggactgct ctccaaggaa ctggacctgt ccagacagtt acactccaag gtcattggag 1620
agaacttctg tatgggcaag cctgagaggg agaggaaaca aaagctgtgt cctggcagaa 1680
ggtctgggtt tgcagatggg tgccctgaat ggaactactt taactaatcc atagggactt 1740
ctggtatgct ttcctctctt tttaaaggaa cttcgtgaca ctaaacatta gcccaaagga 1800
cttcttagcc ttcaattggg agataccttt ggtctgctcc tgcaccaaag ccatatgggt 1860
ggaagtcagt tggcctccct ggttctgcag agggccagaa gaatgagaga gaggaagact 1920
gctggcaggg aaatcgagga ggcgagacta gaactgcacc agcttccctg atgtctgcag 1980
ccatggcttt gcagcgcaga cagagcttct ctgggatgct gggattcttg cctgtatgaa 2040
tgcatcaagt attcatttat tgcccgaata ggcattgcat taagtcctct gtaaggtgtc 2100
aggcaagcc 2109

<210> 15
<211> 1100
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 196774.3.dec

<220>
<221> unsure
<222> 1089
<223> a, t, c, g, or other

<400> 15
ggggtgacct ggcttcagac agctctgccc ttgacctggg gcaaatcact tccctgtgtg 60
ggccttggtt ttctcattcg tgaatgacca gatcactaga gccctggact ctgactttgg 120
gtctccttct tagtttttca caggggaagc tattgtgggt tggtccctac cccacagggc 180
ctgaggcaat ctaacctccc tgagagggtc cctgcagcca gttgcctgag gctgagttga 240
tgtgtgggac agcccaggag gttcctgggg gtggttagtc tgtattcagg gtttggaaga 300
gctgaagtga agtgggcggg gaagtggggg agaggggtgc agttctgcag agaaacgtgg 360
gtgggtagca gagggaccta gaggctgcta gtccaacctt ctgagctctg ggcctttaac 420
tgaacacaga tccttgagga tatggcacta atggagattt gggggctaac tccaaacccc 480
tcactcacaa agggagatgg aggtccagat agggcaccct aagtcacaca gaaccaggcc 540
tcctgcaccc cgttcattgc taatcccata gcactgggct atgagccctt tgggactggg 600
agtctccatg gagtccaacc aagccttcac agggcagggg tgggagggaa ggggctcagg 660
ctgagtgggt ttgtgtctcg cagtttccca gacagtcctg gcccagctgg atgcactgct 720
ggtcttccca ggccaagtgg ctcaactctc ctgcacgctc agcccccagc acgtcaccat 780
cagggactac ggtgtgtcct ggtaccagca gcgggcaggc agtgccctc gatatctcct 840
ctactaccgc tcggaggagg atcaccaccg gcctgctgac atccccgatc gattctcggc 900
agccaaggat gaggcccaca atgcctgtgt cctcaccatt agtcccgtgc agcctgaaga 960
cgacgcggat tactactgct ctgttggcta cggctttagt ccctaggggt ggggtgtgag 1020
atgggtgcct cccctctgcc tcccatttct gcccctgacc ttgggtccct tttaaacttt 1080
ctctgagcnt tgcttccccc 1100

<210> 16
<211> 906
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 233624.11.dec

<220>
<221> unsure
<222> 75, 585
<223> a, t, c, g, or other

<400> 16
cgcttgtgga gctggtggcg gcgctccgca ggggctcggc tgttttccgc gcggcaggcg 60
cggccatggc gcaantggga aagctgctca aggagcagaa gtacgaccgg cagctgaggc 120
gattcctcca tacatttgac agctgtctgg gcccttatgg agcatacagt ggagtggaaa 180
gatggatgct cagcaaacaa aaacaaatga agccaggttg tggggtgatc atgggcaaga 240
ggctttagaa tctgctcatg tttgcctaat aaatgcaaca gccacaggaa ctgaaattct 300
taaaaacttg gtactaccag gtattggttc gtttacaatt attgatggaa atcaggtcag 360
cggagaagat gctggaaaca atttcttcct tcaaagaagc aagtatcggc aagaaccgag 420
ctgaagctgc catggaattc ttacaagaat taaatagcga tgtctctgga agttttgtgg 480
aagagagtcc agaaaacctt ctagacaatg atccctcatt tttctgtagg tttactgttg 540
tagttgcaac tcagcttcct gaaagcactt cactacgctt agcanatgtc ctctggaatt 600
cccagattcc tcttttgatc tgtaggacat atggactagt tggttatatg aggatcatta 660
taaaagaaca tccagtaata gaatctcatc cagataatgc attagaggat ctacgactag 720
ataagccatt tcctgaactg agagaacatt ttcagtccta tgatttggat catatggaaa 780
aaaaggacca cagtcatact ccatggattg tgatcatagc taaatattta gcacagtggt 840
atagtgaaga attctaaaaa atgaaaatgg ggctccagaa gatgaagaga attttgaaga 900
agctat 906

<210> 17
<211> 5923
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 228585.3.dec

<220>
<221> unsure
<222> 3280
<223> a, t, c, g, or other

<400> 17
tcatcagcga tgacagtggg gtcggcgctg aagcactctg ggaccaggtc accatggacg 60
accaggagct ggctttcaaa gctggggacg tcatcgaagt gatggatgcc accaacagag 120
agtggtggtg gggccgggtc gccgatggcg agggctggtt tccagccagc ttcgttcggc 180
tgagggtgaa tcaggacgag cccgcggatg acgacgcccc tctggccggg aacagcggag 240
cggaggacgg cggggcggag gcgcagagca gcaaggacca gatgcggacc aacgtcatca 300
acgagatcct cagcactgag cgggactaca tcaagcacct gcgcgacatc tgcgagggct 360
acgtccggca gtgccgcaag cgcgcagaca tgttcagcga ggagcagctg cgtaccatct 420
tcgggaacat cgaggacatc taccgctgcc agaaggcctt cgtgaaggcc ctggagcaga 480
ggttcaaccg cgagcgccca cacctgagcg agctgggtgc ctgcttcctg gagcatcaag 540
ccgacttcca gatctactcg gagtactgca ataaccaccc caacgcctgc gtggagctct 600
cccggctcac caagctcagc aagtacgtgt acttcttcga ggcctgccgg ctgctgcaga 660
agatgattga catctccctg gatggcttcc tgctgactcc ggtgcagaag atctgcaagt 720
accctctgca gctggccgag ctgctcaaat acacgcaccc ccagcacagg gacttcaagg 780
atgttgaagc cgccttgcat gccatgaaga acgtggccca gctcatcaac gagcggaagc 840
ggagacttga gaacatcgac aagattgctc agtggcagag ctccatagag gactgggagg 900
gagaagatct cttggtcagg agctcagaac tcatctactc gggggagctg actcgagtta 960
cacagcctca agccaaaagc cagcagcgaa tgttctttct ctttgaccac cagctcatct 1020
actgtaagaa ggacctgctc cgccgcgacg tgttgtacta caagggccgg ctggacatgg 1080
acggcctgga ggtggtggac ctggaggacg ggaaggacag agacctccat gtgagcatca 1140
agaacgcctt ccggctgcac cgtggcgcca caggggacag ccacctgctg tgcaccagga 1200
agcccgagca gaagcagcgc tggctcaagg cctttgccag ggagagggag caggtgcagc 1260
tggaccagga gacaggcttc tccatcactg aactgcagag gaagcaggcc atgctgaatg 1320
ccagcaagca gcaggtcaca gggaagccca aagctgttgg ccggccctgc tacctgacgc 1380
gccagaagca cccagccctg cccagcaacc ggccccagca gcaggtcctg gtgctggcgg 1440
agcccagggc gcaagccatt ctaccttctg gcacagcatc agccggctgg caccttccg 1500
caagtgaact ggtccctgcc tgacagcacc tgctgggcct tcctgccagt ggcccccagt 1560

ttttcttccc cgaggcccac tcggcctggc cttcctctgc ctgcaagtga gcagggatgg 1620
gctggggagt tgcttgtgcc accaagacgt gccaggtctg tactcctgtt gtctttttcc 1680
ctgctcctgg tgccctgaag agaccagcaa gggggcagac cccgcactcg ccacaccgcc 1740
gctgcagctt gggccccatc cgccctctgg acctgtgtag ggcctcactg ctggagcggg 1800
gaaaccgcag ctcagcccag gcccagctgg ggagaaggcg ctacctgcgt gggaccctct 1860
tctctggaaa cctaatcctc ctttcatttc ctctgggcag gactctctgg ccttctgtgg 1920
cctgcaatgc caggccatgt gcccctctgc cctctagttc tccaagtccc cagcccggcc 1980
agtggtgcca ggcagcttgc cacttgggag ggcagaagcc aggaattcca caccttgtg 2040
ttgcgcccgg agcccgccct tcgcctccca gcccctcaag acaccgctgg ctgctggaca 2100
ccctcttcac ttgtgtgtgt gtgtgtagcg gaaaaggaca agacggtgca gtcggctgca 2160
tactcccagt cgggagtgtg gtcagtctgc ctgctgctga atcctggggg ctccacccca 2220
gctcgccagg ccctggcttt gctcctggcg ccccttggca ggacagggcg ccatctccac 2280
acacccgctg cctgggctgg gggtcaatcc tgtgtgctga gccacaaaat tcggtctctc 2340
tcttatggct tctcacgctc gtgagcgtaa ggcaatcttc tgtgtcacta aaaatcaatt 2400
ctttttctcc attggttggt ggtagaaaaa caagatgcca aaatccaaac aaaaccagga 2460
acgaggtggt tctggaacta accgcacagc agcaggcaga ctgaccacac tcccgactgg 2520
gagtatgcag ccgactgcac cgtcttgtcc ttctccgcta cacacacaca cacaagtgaa 2580
gagggtgtcc agcagccagc ggtgtcttga ggggctggga ggcgaagggc gggctccggg 2640
cgcaacacaa gggtgtggaa ttcctggctt ctgccctccc aagtggcaag ctgcctggca 2700
ccactggccg ggctggggac ttggagaact agagggcaga ggggcacatg gcctggcatt 2760
gcaggccaca gaaggccaga gagtcctgcc cagaggaaat gaaaggagga ttaggtttcc 2820
agagaagagg gtcccacgca ggtagcgcct tctccccagc tgggcctggg ctgagctgcg 2880
gtttccccgc tccagcagtg aggccctaca caggtccaga gggcggatgg ggcccaagct 2940
gcagcggcgg tgtggcgagt gcggggtctg ccccccttgct ggtctcttca gggcaccagg 3000
agcagggaaa aagacaacag gagtacagac ctggcacgtc ttggtggcac aagcaactcc 3060
ccagcccatc cctgctcact tgcaggcaga ggaaggccag gccgagtggg cctcggggaa 3120
gaaaaactgg gggccactgg caggaaggcc cagcaggtgc tgtcaggcag ggaccattca 3180
cttgcggaag ggtgccagcc ggctgatgct gtgccagaag gtagatgctt gcgcctggct 3240
ccgccagcac caggacctgc tgctggggcc ggttgctggn cagggctggg tgcttctggc 3300
gcgtcaggta gcagggccgg ccaacagctg aacagggaga gacagcagag ggtcatgggt 3360
gcccagggca gagtgccacc ccagctgttg gggacagcat ggggtccgtc ccagggctct 3420
gcttcagagc cgccgtagga ggcagttttg ccacacagga cttgctcctg gagagctgac 3480
cacatccaag agctcaggct gcctggggtc tctggcctgg gcctggcctc ctgcagcagc 3540
ctgaagaggg gggtggcctg tgcaaaggca tgtgatgttt caggggcatg gctctctgta 3600
gaaatgggga aagcagctcc atccctacac caaccccaga tgcctgtcag gaaggaaccc 3660
tggggaaaca agcattgtca ccaatggccc tgatctgggg gagccaaggt gttgagagta 3720
agaatgtcaa ccatatactg acccaggctt ccccagtggc cccttattac caacctttgg 3780
gaacaaacag ctgccaaagg gcagtgggcc cccctcaggc ctggtggcta gctggcttgg 3840
ctttgtggtt cagagaaggg aatgatgtca gcagggtaag gaccgggcag ccgaggaggt 3900
ggggctgctg tccgcctacc tttgggcttc cctgtgacct gctgcttgct ggcattcagc 3960
atggcctgct tcctctgcag ttcagtgatg gagaagcctg tctcctggtc cagctgcacc 4020
tgctccctct ccctggcaaa ggccttgagc cagcgctgct tctgctcagg cttcctggtg 4080
cacagcaggt ggctgtcccc tgtggcgcca cggtgcagcc ggaaggcgtt cttgatgctc 4140
acatggaggt ctctgtcctt cccgtcctcc aggtccacca cctccaggcc gtccatgtcc 4200
agccggccct tgtagtacaa cacgtcgcgg cggagcaggt ccttcttaca gtagatgagc 4260
tggtggtcaa agagaaagaa cattcgctgc tggcttttgg cttgaggctg tgtaactcga 4320
gtcagctccc ccgagtagat gagttctgag ctcctgacca agagatcttc tccctcccag 4380
tcctctatgg agctctgcca ctgagcaatc ttgtcgatgt tctcaagtct ccgcttccgc 4440
tcgttgatga gctgggccac gttcttcatg gcatgcaagg cggcttcaac atccttgaag 4500
tccctgtgct gggggtgcgt gtatttgagc agctcggcca gctgcagagg gtacttgcag 4560
atcttctgca ccggagtcag caggaagcca tccagggaga tgtcaatcat cttctgcagc 4620
agccggcagg cctcgaagaa gtacacgtac ttgctgagct tggtgagccg ggagagctcc 4680
acgcaggcgt tggggtggtt attgcagtac tccgagtaga tctggaagtc ggcttgatgc 4740
tccaggaagc aggcacccag ctcgctcagg tgtgggcgct cgcggttgaa cctctgctcc 4800
agggccttca cgaaggcctt ctggcagcgg tagatgtcct cgatgttccc gaagatggta 4860
cgcagctgct cctcgctgaa catgtctgcg cgcttgcggc actgccggac gtagccctcg 4920
cagatgtcgc gcaggtgctt gatgtagtcc cgctcagtgc tgaggatctc gttgatgacg 4980
ttggtccgca tctggtcctt gctgctctgc gcctccgccc cgccgtcctc cgctccgctg 5040
ttcccggcca gaggggcgtc gtcatccgcg ggctcgtcct gattcaccct cagccgaacg 5100
aagctggctg gaaaccagcc ctcgccatcg gcgacccggc cccaccacca ctctctgttg 5160
gtggcatcca tcacttcgat gacgtcccca gctttgaagc ccagctcctg gtcgtccatg 5220
gtgacatggt cccagagtgc ttcagcgcag accacactgc catcgctgat gagctcattg 5280
atagccagct gctccccacc ccctccaggg tggctgtagt ggtggctgga gctgtgcagg 5340
tcatcataca ggtcctcctc gctccccact tcgtcagcgc agacagctgt gtccagagct 5400
ccatcaggca tggcagtgcc tggtgtgtgc tctggccagc ccatgtggtt cagtcccgtt 5460
ggagcactct gggagagtgg atggcttctt ctagggagat tcaaagactc tggagaggta 5520
atgatcgtct ttctccagca cttgactcca tctttgtgca ctgcccctat gtgcagcctc 5580

ctttcgacgt gggcctgctt ctgcaacttc tttctctgtg tcttctccag ccacaggtca 5640
tccacactga cgggcgagcg gtgagggatg cccctaggct gtgaaggagc ctccccacgc 5700
acagtctctg cagagatgca gaggagggcc catgtcacca tgtcagtggt gaagcagggc 5760
ttctgggcag gctccatgtg gaacgccttc tgactgtgag agcaactggg cttctcacct 5820
gctggttctt cccagggcat gtggttctca ctcccaggaa ggtccctgag ccagggcact 5880
gtgcccaagc cgcgcggacc ctggcctcct tccctgctct ctt 5923

<210> 18
<211> 1228
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 198840.3.dec

<400> 18
gccggcgagt gtgagcggcg cctgctcagg gtagatagct gagggcgggg gtggatgttg 60
gatggattag aaccatcaca cttgggcccg ctgtttgcct gaggttgaac cacaccccga 120
tgtggtgtaa aggaattcat tagccatgga tgtattcatg aaaggacttt caaaggccaa 180
ggagggagtt gtggctgctg ctgagaaaac caaacagggt gtggcagaag cagcaggaaa 240
gacaaaagag ggtgttctct atgtaggctc caaaaccaag gagggagtgg tgcatggtgt 300
ggcaacagtg gctgagaaga ccaaagagca agtgacaaat gttggaggag cagtggtgac 360
gggtgtgaca gcagtagccc agaagacagt ggagggagca gggagcattg cagcagccac 420
tggctttgtc aaaaaggacc agttgggcaa ggaagggtat caagactacg aacctgaagc 480
ctaagaaata tctttgctcc cagtttcttg agatctgctg acagatgttc catcctgtac 540
aagtgctcag ttccaatgtg cccagtcatg acatttctca aagttttac agtgtatctc 600
gaagtcttcc atcagcagtg attgaagtat ctgtacctgc ccccactcag catttcggtg 660
cttccctttc actgaagtga atacatggta gcagggtctt tgtgtgctgt ggattttgtg 720
gcttcaatct acgatgttaa aacaaattaa aaacacctaa gtgactacca cttatttcta 780
aatcctcact attttttgt tgctgttgtt cagaagttgt tagtgatttg ctatcatata 840
ttataagatt tttaggtgtc ttttaatgat actgtctaag aataatgacg tattgtgaaa 900
tttgttaata tatataatac ttaaaaatat gtgagcatga aactatgcac ctataaatac 960
taaatatgaa attttaccat tttgcgatgt gttttattca cttgtgtttg tatataaatg 1020
gtgagaatta aaataaaacg ttatctcatt gcaaaaatat tttatttta tcccatctca 1080
ctttaataat aaaaatcatg cttataagca acatgaatta agaactgaca caaaggacaa 1140
aaatataaag ttattaatag ccatttgaag aaggaggaat tttagaagag gtagagaaaa 1200
tggaacatta accctacact cggaattc 1228

<210> 19
<211> 594
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 082154.5.dec

<400> 19
gtgctgtctt cagaaaagcg aggaccgggc ccgggaggtg aaagttacac aagaactgaa 60
aaacattcaa gttgagcaga tgacaaaact tcaagccaaa catcaagcag aatgtgattt 120
gcttgaagat atgaggacat tcagtcagaa gaaggctgct attgaaagag agtatgcaca 180
gggtatgcag aagttggcta gtcaatacct gaagagagat tggcctggag taaaagctga 240
tgatcggaat gattacagga gcatgtatcc cgtttggaaa tcttttctcg agggaacaat 300
gcaggtagcc cagtctcgga tgaatatatg tgaaaactat aaaaacttca tttctgagcc 360
tgcaaggaca gtgagaagct taaaagaaca gcaactaaaa aggtgtgtgg accagttgac 420
aaagatccaa actgaattac aagagacagt gaaagattta gctaaaggca aaaagaaata 480
ctttgagact gaacagatgg ctcatgcagt acgagagaaa gctgacatcg aggcaaaatc 540
taaacttagt ctttttcaat caagaatcag tttacagaag gcaagtgtaa agtt 594

<210> 20
<211> 4447
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 368396.5.dec

<220>
<221> unsure
<222> 23, 189, 307, 412, 414, 433, 435
<223> a, t, c, g, or other

<400> 20
ctgaggcggc gccggacgga gcnctgcagc ggctgtgaca ggctacgcaa caggttcgcg 60
ggcggcggcc tgacgaccaa gccagctgca gtggcggcga cggcggcaga gcagggtctc 120
cccgcgcctg cccgcgccca ggctgccggt gctgagggac gcggagtcgc gctgtgacgt 180
gcgggaggng cggcgagggc gccagatggc tgagagctag caaggaaaac tcaggaccat 240
gatggctcag tttcccacag ctatgaatgg agggccaaac atgtgggcta ttacctctga 300
agaacgnact aagcatgaca ggcagttttg ataacctcaa accttcagga ggttacataa 360
caggtgatca agcacgtaat tttttcctac aatcaggtct gccggcccct gntnaagctg 420
aactatgggg ttnancagac ctaaacaagg atgggaagat ggatcagcaa gagttctcca 480
tagctatgaa aactcattca aactgaagct tacaaggcca acagttgcct gtggttctcc 540
ctcctattat gaagcaaccc cctatgtttt ctccattaat ttctgctcgt tttggaatgg 600
gaagcatgcc caatctgtcc attcctcagc cattgcctcc agctgcacct ataacatcat 660
tgtcttctgc gacttcaggg accaaccttc ctcccttaat gatgcccact cccctagtgc 720
cttctgttag cacatcatca ttaccaaatg gaaccgccag tctcattcag cctttaccca 780
ttccttattc ttcttcaaca ttgcctcatg ggtcatctta tagtctgatg atgggaggat 840
ttggaggtgc tagtatacag aaagcgcagt ctctgattga tttaggatct agtagctcaa 900
cttcctcgac tgcttcactc tcagggaact cacccaagac tgggacctca gagtgggcag 960
ttcctcagcc tacaagatta aaatatcggc aaaaatttaa tactcttgac aaaagtatga 1020
gtggatatct ctcaggtttt caagctagaa atgcccttct tcagtcaaat ctttctcaaa 1080
ctcagctggc tactatttgg actctggctg acgttgatgg tgatggacag ctaaaagcag 1140
aagagtttat tcttgcaatg caccttactg acatggccaa agctggacag ccattaccac 1200
tgactttacc tcctgagctt gttcctccat ctttcagagg aggaaagcaa attgattcca 1260
ttaatggaac tctgccttca tatcagaaaa tgcaagaaga ggagcctcag aagaaattac 1320
cagttacttt tgaggacaaa cggaaagcca actatgagcg agggaacatg gagctggaaa 1380
agcgacgcca agccttgatg gagcagcaac aaagggaggc agaacgtaaa gcccagaaag 1440
aaaaggaaga gtgggaacga aaacagagag aattacaaga acaagaatgg aagaaacaac 1500
ttgaattaga aaaacgctta gagaagcaac gggaattgga gagacaacga gaggaagaaa 1560
ggagaaaaga catagaaaga cgagaggcag caaaacagga acttgaacga caacgtcgct 1620
tagaatggga gagaattcgg cgacaggagc ttctcaatca aaagaataga gaacaagaag 1680
aaattgtcag gttaaactct aaaaagaaga atcttcatct tgagttggaa gcactgaatg 1740
gcaaacatca gcagatctca ggcagacttc aggatgtccg actcaaaaag caaactcaaa 1800
agactgagct ggaagttctg gataagcagt gtgacttgga aattatggaa atcaagcaac 1860
ttcaacagga acttcaggaa tatcagaata agcttatcta tctggtacct gagaagcaat 1920
tattaaatga aagaattaaa aacatgcagt tcagtaacac acctgattca ggggtcagtt 1980
tacttcataa aaaatcatta gaaaaggaag aattatgcca aagacttaaa gaacagttag 2040
atgctcttga aaaagaaact gcatctaagc tgtcagaaat ggattctttt aacaatcaac 2100
taaaggaact gagagaaacc tacaacacac agcagttagc ccttgaacag ctttataaga 2160
tcaaacgtga caagttgaag gaaattgaaa ggaaaagatt agaactaatg cagaaaaaga 2220
aactagaaga tgagggctgc aaggaaagca aagcaaggaa aagaaaactt atggaaagaa 2280
aatcttagaa aggaggaaga agaaaaacaa aagcgactcc aggaagaaaa aacacaagaa 2340
aaaattcaag aagaggaacg gaaagctgag gagaaacaac gtgagacagc tagtgttttg 2400
gtgaattata gagcattata ccccctttgaa gcaaggaacc atgatgagat gagttttaat 2460
tctggagata taattcaggt tgatgaaaaa accgtaggag aacctggttg gctttatggt 2520
agttttcaag gaaattttgg ctggtttcca tgcaattatg tagaaaaaat gccatcaagt 2580
gaaaatgaaa aagctgtatc tccaaagaag gccttacttc ctcctacagt ttctttatct 2640
gctacctcaa cttcctctga accactttct tcaaatcaac cagcatcagt gactgattat 2700
caaaatgtat ctttttcaaa cctaactgta aatacatcat ggcagaaaaa atcagccttc 2760
actcgaactg tgtcccctgg atctgtatca cctattcatg gacagggaca agtggtagaa 2820
aacttaaaag cacaggccct ttgttcctgg actgcaaaga aagataacca cttgaacttc 2880
tcaaaacatg acattattac tgtcttggag cagcaagaaa attggtggtt tggggaggtg 2940
catggaggaa gaggatggtt tcccaaatct tatgtcaaga tcattcctgg gagtgaagta 3000
aaacgggaag aaccagaagc tttgtatgca gctgtaaata agaaacctac ctcggcagcc 3060
tattcagttg gagaagaata tattgcactt tatccatatt caagtgtgga acctggagat 3120
ttgactttca cagaaggtga agaaatattg gtgacccaga aagatggaga gtggtggaca 3180
ggaagtattg gagatagaag tggaattttt ccatcaaact atgtcaaacc aaaggatcaa 3240
gagagttttg ggagtgctag caagtctgga gcatcaaata aaaaacctga gattgctcag 3300
gtaacttcag catatgttgc ttctggttct gaacaactta gccttgcacc aggacagtta 3360
atattaattc taaagaaaaa tacaagtggg tggtggcaag gagagttaca ggccagagga 3420
aaaaagcgac agaaaggatg gtttcctgcc agtcatgtta aacttttggg tccaagtagt 3480
gaaagagcca cacctgcctt tcatcctgta tgtcaggtga ttgctatgta tgactatgca 3540

gcaaataatg aagatgagct cagtttctcc aagggacaac tcattaatgt tatgaacaaa 3600
gatgatcctg attggtggca aggagagatc aacggggtga ctggtctctt tccttcaaac 3660
tacgttaaga tgacgacaga ctcagatcca agtcaacagt ggtgtgctga tctgcaaacc 3720
ctggacacaa tgcagccaat tgagaggaaa agacagggct atattcatga gctgattcag 3780
accgaagagc ggtacatggc tgaccttcag ctcgtcgtcg aggtgaggag gctgctgctg 3840
gctagctctc ggggtatctg ctgtctctca tgagagatgg tgggcatcag actcagggct 3900
gcctccacgc agagtcaaag caaggcatca cttttgatgt gtgaattcac aaatagtgac 3960
ggaagctctc acatcctcca aatgctgtct ctgcctgccg gataatgctt gagattgaaa 4020
gtctctaatg agctctttcc cccagatgag gtcactcaga gtgaagcggg agaacaagag 4080
ggcattcgca tggcttcgtt ggatgtggca ggagccccat caaggaagga cggggataga 4140
gtggatggga agggcctatg gcagaccgta gcttccttgg atatttgcct aatatctgtt 4200
ttaacatctg acattttcat aaactggtat ctctggagga actgtgaaac agtgaaagtg 4260
ttcacctcat ggttgtatca gtttggaaaa cccaatggga gcatattgta aaatagttcc 4320
caaatatcat atagctactg tttgtttata ccagtgacct ctacgctgat gacagctatc 4380
cttattcgag tagcacttta aaatgatttg tgcttgagtg aacaaaagaa gactttccat 4440
ttctact 4447

<210> 21
<211> 4204
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 349415.4.dec

<400> 21
acgcaggcag tgatgtcacc cagaccacac cccttccccc aatgccactt caggggtac 60
tcagagtcag agacttggtc tgaggggagc agaagcaatc tgcagaggat ggcggtccag 120
gctcagccag gcatcaactt caggaccctg agggatgacc gaaggccccg cccacccacc 180
cccaactccc ccgaccccac caggatctac agcctcagga cccccgtccc aatccttacc 240
ccttgcccca tcaccatctt catgcttacc tccaccccca tccgatcccc atccaggcag 300
aatccagttc cacccctgcc cggaacccag ggtagtaccg ttgccaggat gtgacgccac 360
tgacttgcgc attggaggtc agaagaccgc gagattctcg ccctgagcaa cgagcgacgg 420
cctgacgtcg gcggagggaa gccggcccag gctcggtgag gaggcaaggt aagacgctga 480
gggaggactg aggcgggcct cacctcagac agagggcctc aaataatcca gtgctgcctc 540
tgctgccggg cctgggccac cccgcagggg aagacttcca ggctgggtcg ccactacctc 600
accccgccga cccccgccgc tttagccacg gggaactctg gggacagagc ttaatgtggc 660
cagggcaggg ctggttagaa gaggtcaggg cccacgctgt ggcaggaatc aaggtcagga 720
ccccgagagg gaactgaggg cagcctaacc accaccctca ccaccattcc cgtcccccaa 780
cacccaaccc cacccccatc ccccattccc atccccaccc ccacccctat cctggcagaa 840
tccgggcttt gcccctggta tcaagtcacg gaagctccgg gaatggcggc caggcacgtg 900
agtcctgagg ttcacatcta cggctaaggg agggaagggg ttcggtatcg cgagtatggc 960
cgttgggagg cagcgaaagg gcccaggcct cctggaagac agtggagtcc tgaggggacc 1020
cagcatgcca ggacaggggg cccactgtac ccctgtctca aaccgaggca ccttttcatt 1080
cggctacggg aatcctaggg atgcagaccc acttcagcag ggggttgggg cccagccctg 1140
cgaggagtca tggggaggaa gaagagggag gactgagggg accttggagt ccagatcagt 1200
ggcaaccttg ggctggggga tgctgggcac agtggccaaa tgtgctctgt gctcattgcg 1260
ccttcagggt gaccagagag ttgagggctg tggtctgaag agtgggactt caggtcagca 1320
gagggaggaa tcccaggatc tgcagggccc aaggtgtacc cccaaggggc ccctatgtgg 1380
tggacagatg cagtggtcct aggatctgcc aagcatccag gtgaagagac tgagggagga 1440
ttgagggtac ccctgggaca gaatgcggac tgggggcccc ataaaaatct gccctgctcc 1500
tgctgttacc tcagagagcc tgggcagggc tgtcagctga ggtccctcca ttatcctagg 1560
atcactgatg tcagggaagg ggaagccttg gtctgagggg gctgcactca gggcagtaga 1620
gggaggctct cagaccctac taggagtgga ggtgaggacc aagcagtctc ctcacccagg 1680
gtacatggac ttcaataaat ttggacatct ctcgttgtcc tttccgggag gacctgggaa 1740
tgtatggcca gatgtgggtc ccctcatgtt tttctgtacc atatcaggta tgtgagttct 1800
tgacatgaga gattctcagg ccagcagaag ggagggatta ggccctataa ggagaaaggt 1860
gagggccctg agtgagcaca gaggggatcc tccaccccag tagagtgggg acctcacaga 1920
gtctggccaa ccctcctgac agttctggga atccgtggct gcgtttgctg tctgcacatt 1980
gggggcccgt ggattcctct cccaggaatc aggagctcca ggaacaaggc agtgaggact 2040
tggtctgagg cagtgtcctc aggtcacaga gtagaggggg ctcagatagt gccaacggtg 2100
aaggtttgcc ttggattcaa accaagggcc ccacctgccc cagaacacat ggactccaga 2160
gcgcctggcc tcaccctcaa tactttcagt cctgcagcct cagcatgcgc tggccggatg 2220
taccctgagg tgccctctca cttcctcctt caggttctga ggggacaggc tgacctggag 2280
gaccagaggc ccccggagga gcactgaagg agaagatctg taagtaagcc tttgttagag 2340
cctccaaggt tccattcagt actcagctga ggtctctcac atgctccctc tctccccagg 2400

ccagtgggtc tccattgccc agctcctgcc cacactcccg cctgttgccc tgaccagagt 2460
catcatgcct cttgagcaga ggagtcagca ctgcaagcct gaagaaggcc ttgaggcccg 2520
aggagaggcc ctgggcctgg tgggtgcgca ggctcctgct actgaggagc aggaggctgc 2580
ctcctcctct tctactctag ttgaagtcac cctgggggag gtgcctgctg ccgagtcacc 2640
agatcctccc cagagtcctc agggagcctc cagcctcccc actaccatga actaccctct 2700
ctggagccaa tcctatgagg actccagcaa ccaagaagag gaggggccaa gcaccttccc 2760
tgacctggag tctgagttcc aagcagcact cagtaggaag gtggccaagt tggttcattt 2820
tctgctcctc aagtatcgag ccagggagcc ggtcacaaag gcagaaatgc tggggagtgt 2880
cgtcggaaat tggcagtact tctttcctgt gatcttcagc aaagcttccg attccttgca 2940
gctggtcttt ggcatcgagc tgatggaagt ggaccccatc ggccacgtgt acatctttgc 3000
cacctgcctg ggcctctcct acgatggcct gctgggtgac aatcagatca tgcccaagac 3060
aggcttcctg ataatcatcc tggccataat cgcaaaagag ggcgactgtg cccctgagga 3120
gaaaatctgg gaggagctga gtgtgttaga ggtgtttgag gggagggaag acagtatctt 3180
cggggatccc aagaagctgc tcacccaata tttcgtgcag gaaaactacc tggagtaccg 3240
gcaggtcccc ggcagtgatc ctgcatgcta tgagttcctg tggggtccaa gggccctcat 3300
tgaaaccagc tatgtgaaag tcctgcacca tatggtaaag atcagtggag gacctcgcat 3360
ttcctaccca ctcctgcatg agtgggcttt gagagagggg gaagagtgag tctgagcacg 3420
agttgcagcc agggccagtg ggaggggtc tgggccagtg caccttccgg ggccccatcc 3480
cttagtttcc actgcctcct gtgacgtgag gcccattctt cactctttga agcgagcagt 3540
cagcattctt agtagtgggt ttctgttctg ttggatgact ttgagattat tctttgtttc 3600
ctgttggagt tgttcaaatg ttccttttaa cggatggttg aatgagcgtc agcatccagg 3660
tttatgaatg acagtagtca cacatagtgc tgtttatata gtttaggagt aagagtcttg 3720
ttttttactc aaattgggaa atccattcca ttttgtgaat tgtgacataa taatagcagt 3780
ggtaaaagta tttgcttaaa attgtgagcg aattagcaat aacatacatg agataactca 3840
agaaatcaaa agatagttga ttcttgcctt gtacctcaat ctattctgta aaattaaaca 3900
aatatgcaaa ccaggatttc cttgacttct ttgagaatgc aagcgaaatt aaatctgaat 3960
aaataattct tcctcttcac tggctcgttt cttttccgtt cactcagcat ctgctctgtg 4020
ggaggccctg ggttagtagt ggggatgcta aggtaagcca gactcacgcc tacccatagg 4080
gctgtagagc ctaggacctg cagtcatata attaaggtgg tgagaagtcc tgtaagatgt 4140
agaggaaatg taagagaggg gtgagggtgt ggcgctccgg gtgagagtag tggagtgtca 4200
gtgc 4204

<210> 22
<211> 1044
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 474778.3.dec

<220>
<221> unsure
<222> 231, 979, 984, 999, 1017, 1030-1031, 1037
<223> a, t, c, g, or other

<400> 22
cggaggagcg atctgcaggt ttccatgtca gagcccgatg gagaactgaa gattgccacc 60
tacgcacaaa ggccattgag acacttcgtg tagctggaag acaccaactt cctgacagga 120
gctttatttc atttgggatt tcaagtttac agatggtatc ttctcaaaag ttggaaaaac 180
ctatagagat gggcagtagc gaacccttc ccatcgcaga tggtgacagg nggaggaaga 240
agaagcggag ggccggggcc actgactcct tgccaggaaa gtttgaagat atgtacaagc 300
tgacctctga attgcttgga gagggagcct atgccaaagt tcaaggtgcc gtgagcctac 360
agaatggcaa agagtatgcc gtcaaaatca tcgagaaaca agcagggcac agtcggagta 420
gggtgtttcg agaggtggag acgctgtatc agtgtcaggg aaacaagaac attttggagc 480
tgattgagtt ctttgaagat gacacaaggt tttacttggt ctttgagaaa ttgcaaggag 540
gtacttaccg ttgagtatgt gtgtggactt ctgattaaga cccagggtgg tgatcatcca 600
tcatgaatcc cagagacttc caaaacgagt caagctaata aaaggatgaa ggacttaaaa 660
actgcccttg atttgggaga agggaggccg gagggaagga tgataattag cattttgcag 720
agcttagaat gtcacctgtg tgggtatttt ataaatgcct tttcattata ataggagtca 780
tatatagata catttagtca tcgatttatc aatcccgttt tgacatgttc actgtttaac 840
tacattaatt atggggtgga agctctcaaa tacattgcga agtctagaaa ggctaaaaca 900
gaggagagga gggtcctatt gtttgggaca cagacccagg ataaagggga agcctgagaa 960
tgtgccatcc ttcagatgnt aagntgccat cacccacana cataatcact gcggtgnata 1020
cttaagtgcn ntctganata atga 1044

<210> 23

<211> 3925
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 330933.5.dec

<220>
<221> unsure
<222> 3742, 3746
<223> a, t, c, g, or other

<400> 23
ctggaccagc cgtgcaaatc tctagaagat gacggtgttc tttaaaacgc ttcgaaatca 60
ctggaagaaa actacagctg ggctctgcct gctgacctgg ggaggccatt ggctctatgg 120
aaaacactgt gataacctcc taaggagagc agcctgtcaa gaagctcagg tgtttggcaa 180
tcaactcatt cctcccaatg cacaagtgaa gaaggccact gtttttctca atcctgcagc 240
ttgcaaagga aaagccagga ctctatttga aaaaaatgct gccccgattt tacatttatc 300
tggcatggat gtgactattg ttaagacaga ttatgaggga caagccaaga aactcctgga 360
actgatggaa aacacggatg tgatcattgt tgcaggagga gatgggacac tgcaggaggt 420
tgttactggt gttcttcgac gaacagatga ggctaccttc agtaagattc ccattggatt 480
tatcccactg ggagagacca gtagtttgag tcataccctc tttgccgaaa gtggaaacaa 540
agtccaacat attactgatg ccacacttgc cattgtgaaa ggagagacag ttccacttga 600
tgtcttgcag atcaagggtg aaaaggaaca gcctgtattt gcaatgaccg gccttcgatg 660
gggatctttc agagatgctg gcgtcaaagt tagcaagtac tggtatcttg ggcctctaaa 720
aatcaaagca gcccactttt tcagcactct taaggagtgg cctcagactc atcaagcctc 780
tatctcatac acgggaccta cagagagacc tcccaatgaa ccagaggaga cccctgtaca 840
aaggccttct ttgtacagga gaatattacg aaggcttgcc gtcctactgg gcacaaccac 900
aggatgccct ttcccaagag gtgagcccgg aggtctggaa agatgtgcag ctgtccacca 960
ttgaactgtc catcacaaca cggaataatc agcttgaccc gacaagcaaa gaagattttc 1020
tgaatatctg cattgaacct gacaccatca gcaaaggaga ctttataact ataggaagtc 1080
gaaaggtgag aaaccccaag ctgcacgtgg agggcacgga gtgtctccaa gccagccagt 1140
gcactttgct tatcccggag ggagcagggg gctcttttag cattgacagt gaggagtatg 1200
aagcgatgcc tgtggaggtg aaactgctcc ccaggaagct gcagttcttc tgtgatccta 1260
ggaagagaga acagatgctc acaagcccca cccagtgagc agcagaagac aagcactctg 1320
agaccacact ttaggccacc ggtgggacca aaagggaaca ggtgcctcag ccatcccaac 1380
agtgtcgtca gagggtcccc agggcatttt catggcaagt acccctctgc ccccactcca 1440
gcagtgcttc ccaaagtgtg ctctgtcacc tgctttgcaa tcggcttcca ttagcgcatg 1500
ttttattttg gtgtgacggt tggccctcct aaacacggac tttcctcagg ctggttcaag 1560
acggaaaagg actttcttct gttttcttcc aaagtgcaac cacagtggag agcccacggt 1620
gggcttagcc tgcctaggcc cttccatttc tcttctttga ccgtgctagg aattccagga 1680
aagtgcattc ctgccctggt gacctttttcc tatgtctagg ctcctccaca ggtgctgcta 1740
ttttgtgagc tccggctcct gtttagcttt tatttcagtt ctaacctcag tccagaaaca 1800
tatgtgaggt tgtttccctc ttcagccacg gctacaatac cggaaaatgc tagtttttat 1860
ttatttttt aagtagtgct tcctaaatgg tttgcatgag agccacctgg ggtacatgtt 1920
gaaaacttat ttggggtcta ccccaaacct aataacccaa atttggggat ggggcccagg 1980
aatatgcatt tttaaaaagt catctgccct tcccaggtga ttctgtaagt tgtccctcaa 2040
ctgtacttgg agaaatcgtg ttttaaagca gtagtccaca aagtattctg ctcatgtgcc 2100
cccaaaagta ttttgaaaaa tcatgtatac cctcacccat ctaagttgat atctaaaatt 2160
ttatctaagt tggtatctaa aatttttcat gggaagttaa atagttgaca aagtatgtat 2220
ttgctggtgt cgtgtaaata ttggtatttt aaaataaaaa ctgttacatc actattttaa 2280
acatatccag tacaatttaa atatcacaac aatttgacac ccttcattca tttataaaaa 2340
taaatgagct agttctttag tagttaaaca tttcaaattg gcttttctcc ttctgtattt 2400
ccataccact tttcagccaa gaatcctatc ataatgtaat ctattatgcc cgacatcttt 2460
taatcattca ccccattact tcttgtcaac aaaaaatata aatggaaatt tttttttag 2520
ctcttgcttt aagtgtttgt ttgttatctc agtccagaac caatattatc gtaattaatt 2580
attggtatat aatgaaaacg gtattaattc ttggatgatt aaaagttttt ttattagaat 2640
gttctttatc ctaattagtt catttatcca agaatacatg aatgtgattt acagctgaga 2700
tggggttcaa cctcagctgt attccttgtt tctgtataga tgtaagcaca taaattcgat 2760
ggaatagaat tacgttaaca atgtttttac agttctttgg attcctttgg cattttgaca 2820
aagatcacag tgctctatca tcaagaatta ttaatgatga tctatcaact aacaaacaac 2880
ttgattagat tctcctttag tctgttgaaa gcagagaact gaaatccacc tgatttacca 2940
tggctttgcc agccagtcat tagcaccatt tacttttact atcgctgaca ttttcctttg 3000
ttcagtggcc ctgaggttct tacactctag ggggcagtgc accacaggaa gatagatcaa 3060
tgagggagga ttgcgagggg gaaggggagg aagcagagct ggcaggcctt agctacaggc 3120
tctctctcag gcagatccct tttaagatac atacaccatg cccacacatc ccatggagag 3180

agaccaatgc tttagtagat tacagaacag ctatgaaaag tccatgaatg aagatcacaa 3240
aaaggaaggc tttcttattt catactgtat tcttcagggt ggtaaaattt ctgcttttgg 3300
caaaaacata acagacggtt ccaaacatca gcataaagat cactcatccc ataccaccca 3360
caggtaggga ggaaggatgc tgtagtatat gaaaacaaaa gttttcacct gagctgagag 3420
catttagcat atcgtggttc tgtaacaata tcaaggacca gtgcagaatc tggctttctt 3480
ttctgatagg ctaccagtgt gtgtttatgt gtgctcattt tgtggttcta atcataatgg 3540
tacatataat tagggaagga tatggaagcc actttagaat cttattcatt tttaaatata 3600
aatatgcctt gtttcaaact ttgttttctt gattcaggct ttctttcctg tgagggcttg 3660
gtttccttat tgttgactgc tttgttcttt gccttgtcct tccctataaa gcctgcatgg 3720
aagacgttta atagtgcaac tnaaanagag tcagctgagt gaggcttgtc agccaaagct 3780
ggaatgctga gctatctgga ggagatccta ataacccaat ttggggatgg ggcccaggaa 3840
tctgcattgg gaagtcggcc acccttccca ggtgattgtt tagtacaaac tttttgacag 3900
attaatttca ctcaaatgca aagat 3925

<210> 24
<211> 1254
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 998036.2.dec

<220>
<221> unsure
<222> 36, 49, 61, 63, 77, 85
<223> a, t, c, g, or other

<400> 24
gcggcgctga gccgcgccgc cgccactgag gaagangccg gcccagccnc cgccgcgtcc 60
ngncccgcgc gcctggntcc cagcnccccg atcccggcgc cccaaccccc acgcccgcct 120
ccgccaactt tcacgctgcc tcggcggccc ggcccggctc gacgccaatg gtggaggcca 180
tagtggagtt tgactaccag gcccagcacg atgatgagct gacgatcagc gtgggtgaaa 240
tcatcaccaa catcaggaag gaggatggag gctggtggga gggacagatc aacggcagga 300
gaggtttgtt ccctgacaac tttgtaagag aaataaagaa agagatgaag aaagaccctc 360
tcaccaacaa agctccagaa aagcccctgc acgaatgccc agtggaaact ctttgctgtc 420
ttctgaaacg attttaagaa ccaataagag aggcgagcga cggaggcgcc ggtgccaggt 480
ggcattcagc tacctgcccc agaatgacga tgaacttgag ctgaaagttg gcgacatcat 540
agaggtggta ggagaggtag aggaaggatg gtgggaaggt gttctcaacg ggaagactgg 600
aatgtttcct tccaacttca tcaaggagct gtcaggggag tcggatgagc ttggcatttc 660
ccaggatgag cagctatcca agtcaagttt aagggaaacc acaggctccg agagtgatgg 720
gggtgactca agcagcacca agtctgaagg tgccaacggg acagtggcaa ctgcagcaat 780
ccagcccaag aaagttaagg gagtgggctt tggagacatt ttcaaagaca agccaatcaa 840
actaagacca aggtcaattg aagtagaaaa tgactttctg ccggtagaaa agactattgg 900
gaagaagtta cctgcaacta cagcaactcc agactcatca aaaacagaaa tggacagcag 960
gacaaagagc aaggattact gcaaagtaat atttccatat gaggcacaga atgatgatga 1020
attgacaatc aaagaaggag atatagtcac tctcatcaat aaggactgca tcgacgtagg 1080
ctggtgggaa ggagagctga acggcagacg aggcgtgttc cccgataact tcgtgaagtt 1140
acttccaccg gactttgaaa aggaagggaa tagacccaag aagccaccgc ctccatccgc 1200
tcctgtcatc aaacaagggg caggcaccac tgagagaaaa catgaaatta aaaa 1254

<210> 25
<211> 499
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 999304.1.dec

<400> 25
cggacgcgtg ggctaggccc ccagtgctgt cactctcacc catcctcctc tacacatgtg 60
agatgtttca ggacccagtg gcctttgatg atgttgctgt gaacttcacc caggaggagt 120
gggctttgct ggatatttcc cagaggaaac tctacaagga agtgatgctg gaaactttca 180
ggaacctgac ctctgtagga aaaagttgga aagaccagaa cattgaatat gagtaccaaa 240
accccaggag aaacttcagg agtctcatag aaaagaaagt caatgaaatt aaagatgaca 300
gtcattgtgg agaaactttt acccaggttc cagatgacag gctgaacttc caggagaaga 360

```
aagcttctcc tgaaataaaa tcatgtgaca gctttgtgtg tggagaagtt ggcctaggta 420
actcatcttt taatatgaac atcagaggtg acattggaca caaggcctat gagtatcagg 480
aatatggacc gaagccata                                              499
```

Docket No.: PT-1086 USN

DECLARATION AND POWER OF ATTORNEY FOR UNITED STATES PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name, and

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if more than one name is listed below) of the subject matter which is claimed and for which a United States patent is sought on the invention entitled

MOLECULES FOR DISEASE DETECTION AND TREATMENT

the specification of which:

/___/ is attached hereto.

/___/ was filed on _____ as application Serial No. _____ and if this box contains an X /___/, was amended on _____.

/ X / was filed as Patent Cooperation Treaty international application No. PCT/US00/26085 on September 22, 2000, if this box contains an X /_/, was amended on under Patent Cooperation Treaty Article 19 on _____ 2001, and if this box contains an X /_/, was amended on _____ _____.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge my duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim the benefit under Title 35, United States Code, §119 or §365(a)-(b) of any foreign application(s) for patent or inventor's certificate indicated below and of any Patent Cooperation Treaty international applications(s) designating at least one country other than the United States indicated below and have also identified below any foreign application(s) for patent or inventor's certificate and Patent Cooperation Treaty international application(s) designating at least one country other than the United States for the same subject matter and having a filing date before that of the application for said subject matter the priority of which is claimed:

Country	Number	Filing Date	Priority Claimed
_____	_____	_____	/_/ Yes /_/ No
_____	_____	_____	/_/ Yes /_/ No

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)
60/156,565	September 28, 1999	Expired
60/168,197	November 30, 1999	Expired

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of Title 35, United States Code §112, I acknowledge my duty to disclose material information as defined in Title 37 Code of Federal Regulations, §1.56(a) which occurred between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)

I hereby appoint the following:

Lucy J. Billings	Reg. No. 36,749
Michael C. Cerrone	Reg. No. 39,132
Diana Hamlet-Cox	Reg. No. 33,302
Richard C. Ekstrom	Reg. No. 37,027
Barrie D. Greene	Reg. No. 46,740
Matthew R. Kaser	Reg. No. 44,817
Lynn E. Murry	Reg. No. 42,918
Shirley A. Recipon	Reg. No. 47,016
Susan K. Sather	Reg. No. 44,316
Michelle M. Stempien	Reg. No. 41,327
David G. Streeter	Reg. No. 43,168
Stephen Todd	Reg. No. 47,139
Christopher Turner	Reg. No. 45,167
P. Ben Wang	Reg. No. 41,420

respectively and individually, as my patent attorneys and/or agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith. Please address all communications to:

LEGAL DEPARTMENT
INCYTE GENOMICS, INC.
3160 PORTER DRIVE, PALO ALTO, CA 94304

TEL: 650-855-0555 FAX: 650-849-8886 or 650-845-4166

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

First Joint Inventor:

Full name: David M. Hodgson

Signature:

Date: FEB 13 , 2001

Citizenship ~~United States of America~~ UNITED KINGDOM

Residence: ~~Palo Alto, California~~ ANN ARBOR, MICHIGAN

P.O. Address: 2795 WINDWOOD DR., APT. 1105
~~567 Addison Avenue~~
~~Palo Alto, California 94301~~
ANN ARBOR, MICHIGAN 48105

Second Joint Inventor:

Full name: Stephen E. Lincoln

Signature:

Date: Feb 1 , 2001

Citizenship United States of America

Residence: Redwood City, California CA

P.O. Address: 725 Sapphire Street
Redwood City, California 94061

Third Joint Inventor: 3-00

Full name: Frank D. Russo

Signature:

Date: 16 Feb , 2001

Citizenship United States of America

Residence: Sunnyvale, California CA

P.O. Address: 939 Rosette Court
Sunnyvale, California 94086

Fourth Joint Inventor: 4-00

Full name: Peter A. Spiro

Signature:

Date: Mar 6 , 2001

Citizenship United States of America

Residence: Palo Alto, California CA

P.O. Address: 410 Sheridan Ave. #333
Palo Alto, California 94306

Fifth Joint Inventor: 5-00

Full name: Steven C. Banville

Signature:

Date: Feb 2 , 2001

Citizenship United States of America

Residence: ~~Palo Alto~~, SUNNYVALE, CA California

P.O. Address: ~~365 Monroe Drive~~
~~Palo Alto, California 94306~~
604 SAN DIEGO AVENUE
SUNNYVALE, CALIFORNIA
94085-
3047

2/2/01

Sixth Joint Inventor: 6-0

Full name:	Shawn R. Bratcher
Signature:	Shawn Bratcher
Date:	Feb 2, 2001
Citizenship	United States of America
Residence:	Mountain View, California CA
P.O. Address:	550 Ortega Ave., #B321, Mountain View, California 94040

Seventh Joint Inventor: 7-0

Full name:	Gerard E. Dufour
Signature:	G. E. Dufour
Date:	Feb. 16, 2001
Citizenship	United States of America
Residence:	Castro Valley, California CA
P.O. Address:	5327 Greenridge Rd. Castro Valley, California 94552-2619

Eighth Joint Inventor: 8-0

Full name:	Howard J. Cohen
Signature:	Howard J Cohen
Date:	5 Feb, 2001
Citizenship	United States of America
Residence:	Palo Alto, California CA
P.O. Address:	3272 Cowper Street Palo Alto, California 94306-3004

Ninth Joint Inventor: 9-00

Full name: Bruce H. Rosen

Signature:

Date: Feb 14, 2001

Citizenship United States of America

Residence: Menlo Park, California CA

P.O. Address: 177 Hanna Way
Menlo Park, California 94025

Tenth Joint Inventor: 10-00

Full name: Purvi Shah

Signature: Purvi Shah

Date: Feb 23, 2001

Citizenship India

Residence: San Jose, California CA

P.O. Address: 859 Salt Lake Drive
San Jose, California 95133

Eleventh Joint Inventor: 11-00

Full name: Michael S. Chalup

Signature: Michael S. Chalup

Date: Feb 6, 2001

Citizenship United States of America

Residence: Sunnyvale, California CA

P.O. Address: M.C. 2/6/01
183 Abalones Dr., Apt. 6
Sunnyvale, California 94086

Twelfth Joint Inventor: 12-00

Full name: Jennifer L. Hillman
Signature:
Date: Feb 16, 2001
Citizenship United States of America
Residence: Mountain View, California CA
P.O. Address: 230 Monroe Drive, #17
Mountain View, California
94040

Thirteenth Joint Inventor: 13-00

Full name: Anissa L. Jones
Signature:
Date: 3/25/02, 2001 AJ 3/25/02
Citizenship United States of America
Residence: San Jose, California CA
P.O. Address: 445 South 15th St.
San Jose, California 95112

Fourteenth Joint Inventor: 14-00

Full name: Jimmy Y. Yu
Signature:
Date: Feb 16, 2001
Citizenship United States of America
Residence: Fremont, California CA
P.O. Address: 3655 Wyndham Dr.
Fremont, California 94536

Fifteenth Joint Inventor: 15-00

Full name: Lila B. Greenawalt

Signature: Lila B Greenawalt

Date: FEBRUARY 9, 2001

Citizenship United States of America

Residence: San Jose, California CA

P.O. Address: 1596 Ballantree Way
San Jose, California 95118-2106

Sixteenth Joint Inventor: 16-00

Full name: Scott R. Panzer

Signature: Scott R Panzer

Date: 5 Feb, 2001

Citizenship United States of America

Residence: Sunnyvale, California CA

P.O. Address: 571 Bobolink Circle
Sunnyvale, California 94087

Seventeenth Joint Inventor: 17-00

Full name: Ann M. Roseberry

Signature: Ann M Roseberry

Date: February 5, 2001

Citizenship United States of America

Residence: Redwood City, California CA

P.O. Address: 725 Sapphire Street
Redwood City, California 94061

Eighteenth Joint Inventor: 18-00

Full name:	**Rachel J. Wright**
Signature:	R.J. Wright
Date:	February 14th , 2001
Citizenship	**New Zealand**
Residence:	**Mountain View, California** CA
P.O. Address:	**333 Anna Ave. Mountain View, California 94043**

Nineteenth Joint Inventor: 19-00

Full name:	**Wensheng Chen**
Signature:	
Date:	Feb 2 , 2001
Citizenship	**China**
Residence:	**Mountain View, California** CA
P.O. Address:	**210 Easy Street #25 Mountain View, California 94043**

Twentieth Joint Inventor: 20-00

Full name:	**Tommy F. Liu**
Signature:	Tommy F. Liu
Date:	Feb 16 , 2001
Citizenship	**United States of America**
Residence:	**Daly City, California** CA
P.O. Address:	**201 Ottilia Street Daly City, California 94014**

Twenty-First Joint Inventor: 21-00

Full name: Pierre E. Yap

Signature: Pierre E. Yap

Date: Feb. 13 , 2001

Citizenship United States of America

Residence: Lafayette, California CA

P.O. Address: 201 Happy Hollow Court
Lafayette, California
94549-6243

Twenty-Second Joint Inventor: 22-00

Full name: Theresa K. Stockdreher

Signature: Theresa K. Stockdreher

Date: FEB 08 , 2001

Citizenship United States of America

Residence: Sunnyvale, California CA

P.O. Address: 1596 Ontario Drive, #2
Sunnyvale, California 94087

Twenty-Third Joint Inventor: 23-00

Full name: Stefan Amshey

Signature: Stefan Amshey

Date: February 12 , 2001

Citizenship United States of America

Residence: ~~Mountain View,~~ San Francisco CA California SRA 02/12/2001

P.O. Address: ~~1541 Canna Court~~
~~Mountain View, California~~
~~94043~~ SRA 02/12/2001

1605 20th St
San Francisco, CA 94107

Twenty-Fourth Joint Inventor:

24-0

Full name:	Willy T. Fong
Signature:	Willy T. Fong
Date:	February 12, 2001
Citizenship	United States of America
Residence:	San Francisco, California CA
P.O. Address:	572 Cambridge Street San Francisco, California 94134